Medical Image Analysis 14 (2010) 87–110

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

A review of automatic mass detection and segmentation in mammographic images

Arnau Oliver [a,*], Jordi Freixenet [a], Joan Martí [a], Elsa Pérez [b], Josep Pont [b], Erika R.E. Denton [c], Reyer Zwiggelaar [d]

[a] Dept. of Computer Architecture and Technology, University of Girona Ed. P-IV, Campus de Montilivi 17071, Girona, Spain
[b] Dept. of Radiology, University Hospital Josep Trueta, Avda de França, s/n 17007, Girona, Spain
[c] Dept. of Breast Imaging, Norfolk and Norwich University Hospital, Norwich NR4 7UY, UK
[d] Dept. of Computer Science, University of Wales, Aberystwyth SY23 3DB, UK

ARTICLE INFO

Article history:
Received 30 June 2009
Received in revised form 15 December 2009
Accepted 18 December 2009
Available online 29 December 2009

Keywords:
Digital mammography
Computer-Aided Diagnosis systems
Breast mass detection
Automatic segmentation
FROC analysis

ABSTRACT

The aim of this paper is to review existing approaches to the automatic detection and segmentation of masses in mammographic images, highlighting the key-points and main differences between the used strategies. The key objective is to point out the advantages and disadvantages of the various approaches. In contrast with other reviews which only describe and compare different approaches qualitatively, this review also provides a quantitative comparison. The performance of seven mass detection methods is compared using two different mammographic databases: a public digitised database and a local full-field digital database. The results are given in terms of Receiver Operating Characteristic (ROC) and Free-response Receiver Operating Characteristic (FROC) analysis.

1. Introduction

Breast cancer is considered a major health problem in western countries, and constitutes the most common cancer among women in the European Union (Eurostat, 2002). A recent study developed by the American Cancer Society estimates that, in the United States, between one in eight and one in 12 women will develop breast cancer during their lifetime (American Cancer Society, 2007). Breast cancer remains, in the United States as well in the European Union, the leading cause of death for women after their 40 s (Eurostat, 2002; Buseman et al., 2003). However, although breast cancer incidence has increased over the past decade, breast cancer mortality has declined among women of all ages (Sickles, 1997). This favourable trend in mortality reduction may relate to the widespread adoption of mammography screening (Sickles, 1997; De Koning et al., 1995), and improvements made in breast cancer treatment (Buseman et al., 2003).

Mammography remains the key screening tool for the detection of breast abnormalities. Vacek et al. (2002) show that the proportion of breast tumours that were detected in Vermont by mammographic screening increased from 2% during 1974–1984 to 36% during 1995–1999. However, it is also well known that expert radiologists can miss a significant proportion of abnormalities (Bird et al., 1992; Birdwell et al., 2001). In addition, a large number of mammographic abnormalities turn out to be benign after biopsy (Basset and Gold, 1987; Hall et al., 1988).

There are a large number of different types of mammographic abnormality (Kopans, 1998). In the majority of cases, however, the abnormalities are either micro-calcifications or masses. Micro-calcifications usually form clusters and individual micro-calcifications can range from 20 to several hundred microns in diameter. On the other hand, a breast mass is a generic term to indicate a localised swelling, protuberance, or lump in the breast. Masses can be caused by different processes: from natural changes in the breast to cancerous processes. Masses are characterised by their location, size, shape, margin, and associated findings (i.e. architectural distortion, contrast). Fig. 1 shows different masses according to their shape and margin (the border of the mass). These associated properties are examined by radiologists as they are strongly correlated with the classification (benign versus malignant) of the mass (Heywang-Köbrunner et al., 2001). It is generally accepted that mass detection is a more challenging problem than the detection of micro-calcifications, not only for the large variation in size and shape in which masses can appear in a mammogram but also because masses often exhibit poor image contrast (Basset and Gold, 1987; Vyborny and Giger, 1994).

The idea of developing computer systems to assist radiologists in the detection and classification of breast cancer is not recent (Winsberg et al., 1967). However, the recent development of full-field digital mammographic imaging systems has been a catalyst in the increase of such computer systems (Kuzmiak et al., 2002).

* Corresponding author. Tel.: +34 972 418878; fax: +34 972 418259.
E-mail address: aoliver@eia.udg.edu (A. Oliver).

1361-8415/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.media.2009.12.005

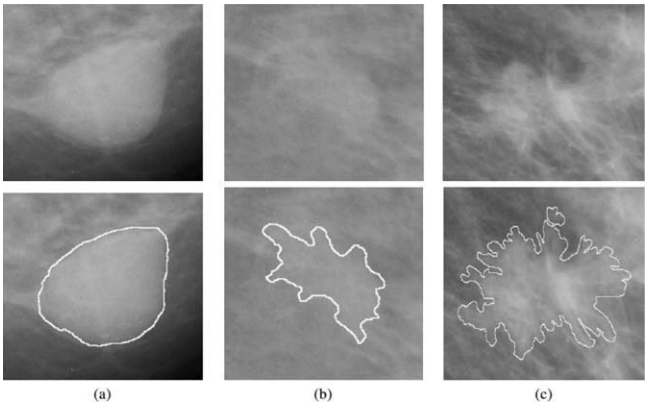

Fig. 1. Three mass examples with different shape and margin: (a) circular shape and circumscribed margin, (b) lobular shape and well defined margin, and (c) spiculated shape and ill-defined margin. The last of the three has a higher malignancy probability.

A Computer-Aided Diagnosis (CAD) system is a set of automatic or semi-automatic tools developed to assist radiologists in the detection and/or classification of mammographic abnormalities (Birdwell et al., 2001; Freer and Ulissey, 2001).

In 2001 Freer and Ulissey (2001), using a database containing 12,860 patients, concluded that the use of CAD in the interpretation of screening mammograms can increase the detection of early-stage malignancies. However, results published in 2005 using a database of 6111 women indicated that in an established screening environment CAD, in its present form, is not effective in that there was no significant difference observed in cancer detection rates with CAD compared to reading mammograms without CAD (Khoo et al., 2005). These are only two examples of a still open debate in the medical imaging field (Gur et al., 2004; Nishikawa and Kallergi, 2006; Fenton et al., 2007). It has been agreed that CAD systems are useful for detecting cancers that are clinically missed. However, the main drawback of such systems is both the significant number of false positive detections (Nishikawa and Kallergi, 2006) and true masses detected only in one view (Khoo et al., 2005; Zheng et al., 2006), which currently reduces their clinical use. For the low number of malignancies within the screening population, which is around 6 out of 1000 screened cases, this is a major issue (Taylor et al., 2005; Astley et al., 1998).

The first part of the review is focused on published techniques related to the detection and segmentation of mammographic masses of any shape, margin and size. Note that we distinguish between detection and segmentation algorithms. Detection is defined as the identification of potential lesions within all the parenchymal background. Usually, these methods generate a marker/prompt at a suspicious region in a mammogram. In contrast, segmentation is defined as a method able to detect the precise outline of the potential lesion. It should be noted that using these definitions, there are algorithms that, at the same time, detect and segment masses. Looking at the literature, there are three possible outputs for mass detection/segmentation algorithms: the detection (and/or segmentation) of potential lesions, the classification of the detected lesion as mass or not mass (usually referred as false positive reduction algorithms), or the diagnosis of a lesion (classification as a benign or malignant mass). Here we concentrate on the detection (and segmentation) stage, which is considered the first and key stage in the complete process. The remaining two stages usually comprise a two class discrimination process (either false positive versus true positive or benign versus malignant).

The second part of the paper is a quantitative review of relevant approaches to mass detection. We re-implemented seven different algorithms, and we evaluated them using two different databases. The aim is not just to compare the performance of the algorithms but mainly to obtain conclusions from the strategies used. In detail, we want to answer the following two questions: does the density of the breast affect the detection of lesions? and do the size and the shape of a lesion modify the performance of the algorithms? Both questions as well as related ones are addressed in the second part of the paper.

This is not the first attempt to review mammographic detection or segmentation. For instance, in 1994, Vyborny and Giger (1994) reviewed different strategies. Recently, Cheng et al. (2003, 2006), Rangayyan et al. (2007), and Elter and Horsch (2009) published surveys in mammographic computer-aided diagnosis. Cheng et al. (2003) covers the detection and classification of micro-calcifications. Cheng et al. (2006) cover the general enhancement of mammographic images, the detection and classification of masses, and underlying computer vision techniques, with an approximate even split in references between these four areas. They provide a review covering a similar research area, but do not provide a distinct classification between the various approaches and concentrate on the specific technical approaches to group publications together: the different approach to the review makes this complementary to the current review. Rangayyan et al. (2007) provided a review covering the full range of mammographic abnormalities using an approach similar to Cheng et al. (2006). The coverage of the detection and classification of masses is limited to a description of some methods and how these are evaluated. Both Cheng et al. (2006) and Rangayyan et al. (2007) qualitatively deal with enhancement, detection, characterisation, and classification of masses. Finally, Elter and Horsch (2009) focused their review on approaches for mass and micro-calcification diagnosis, covering the segmentation of region of interests for extracting shape and contour features, and their posterior classification. In contrast with these reviews, our review is focused only on mass detection and segmentation (all relevant references from Cheng et al. (2006), Rangayyan et al. (2007), and Elter and Horsch (2009) are covered here), highlighting the different strategies used for detecting suspicious regions and grouping publications together according to the used computer vision based methodologies (Fu and Mui, 1981). In addition, a quantitative comparison of seven selected algorithms is given in terms of Free-response Receiver Operating Characteristic (FROC) and Receiver Operating Characteristic (ROC) analysis. In this context, the works by Singh and Al-Mansoori (2000) and te Brake and Karssemeijer (2001) should also be mentioned, as they compared two strategies: region growing and contour-based detection (the former using contour filters over an enhanced image and the latter dynamic programming). Timp and Karssemeijer (2004) added to this last comparison an active contour algorithm.

The remainder of this paper is structured in five sections. In Sections 2 and 3 a review of mass detection and segmentation techniques is presented. Approaches are discussed and classified according to computer vision strategies using the work of Fu and Mui (1981) as a basis. Section 2 reviews approaches that use only a single mammographic view to detect masses, whilst Section 3 explains the different strategies using multiple mammographic views in the detection process. In Section 4, the performance of a subset of representative techniques is evaluated in detail using two different mammographic databases. Discussions are given in Section 5, and in Section 6 the conclusions of this review and suggestions for future research directions are presented.

2. Mass detection using a single view

Detection using a single mammographic image relies on the fact that pixels inside a mass have different characteristics from the other pixels within the breast area. The characteristics used can

be simply related to grey-level intensity values or to (local) texture or morphological measures. In addition, some approaches take the distribution of spicules associated with masses into account. Both aspects can be treated independently or sequentially. The columns of Tables 1–4 classify existing methods according to the characteristics used for the detection/segmentation. Note that the features used in the optional subsequent classification processes (benign/malignant discrimination or false positive reduction) are not taken into account here. In the lasts columns of each table we indicate the aim of each work (which can be the detection of masses in full mammograms, the segmentation of a mass given a small patch of the mammogram or a given seed point, or the detection and segmentation of the mass given the full image) and the size of the database used for validating the approach (in parenthesis the number of images used for training the parameters). Note that depending of the aim of the approach the used images can be regions of interests (ROIs), single mammograms, pairs of mammograms or full four-mammogram cases.

In generic computer vision terminology, segmentation techniques can be divided into unsupervised and supervised approaches. **Supervised segmentation**, or

- **Model-based methods**, rely on the prior knowledge about the object and background regions to be segmented. The prior information is used to determine if specific regions are present within an image or not.

Alternatively, **unsupervised segmentation** partitions an image into a set of regions which are distinct and uniform with respect to specific properties, such as grey-level, texture or colour. Classical approaches to solving unsupervised segmentation are divided in three major groups (Fu and Mui, 1981):

- **Region-based methods**, which divide the image into homogeneous and spatially connected regions.
- **Contour-based methods**, which rely on the boundaries of regions.
- **Clustering methods**, which group together those pixels having the same properties and might result in non-connected regions.

With regard to mass detection, we should clarify that unsupervised approaches tend to be combined with a posterior false positive reduction step in order to increase their performance. This false positive reduction step usually consists of a feature extraction algorithm combined with a classifier. In contrast, the supervised approaches usually combine both steps into a single one.

In subsequent subsections, a more detailed description of mass detection and segmentation methods in each category (region-based, contour-based, clustering and model-based methods) is given.

2.1. Region-based methods

Region-based segmentation relies on the principle of homogeneity, which means that there has to be at least one feature that remains uniform (±an error measure) for all pixels within a region. Table 1 shows the mass detection and/or segmentation approaches which are based on such a principle. The last column of the table indicates the aim of the described approaches. The algorithms labelled as segmentation usually start from a set of seed points manually located, or they are applied over regions of interests, which are small patches of a mammogram containing a suspicious region.

Region-based methods can be split in two basic strategies: the well known *region growing* and *split and merge* approaches. Watershed based methods are related to the former and also discussed here.

2.1.1. Region growing and related methods

More than 30 years have passed since Zucker reviewed region growing algorithms (Zucker, 1976). Region growing is based on the propagation of an initial seed point according to a specific homogeneity criterion, iteratively increasing the size of the region. Since those early days, region growing has seen a number of improvements, mainly due to the integration of boundary information in the algorithm. As Freixenet et al. (2002) reviewed, this information can be introduced before the growing step, using for example a controlled seed placement (Benois and Barba, 1992) or during growing, like in active region algorithms (Zhu and Yuille, 1996).

Table 1
Region-based mammographic detection and/or segmentation techniques indicating the used feature information. The penultimate column indicates the number of images used for validating the approach (r stands for ROI, m for mammogram, p for pair (2 mammograms), and c for case (4 mammograms)) and the parenthesis stand for the number of images used for training the approaches. The last column indicates the purpose of the described approaches: detection of masses in full mammograms (D), the segmentation of a mass given a small patch of the mammogram or a given seed point (S), or the detection and segmentation of the mass given the full image (D and S).

Author	Year	Texture	Gradient	Grey-level	Shape	Num testing	D/S
Bårman and Granlund (1994)	1994		√	√			S
Woods and Bowyer (1994)	1994			√		330 m	D and S
Huo et al. (1995)	1995			√	√	95 r	S
Zheng et al. (1995b, 2003b)	1995			√	√	510(90) m	D and S
Pohlman et al. (1996)	1996			√		96 r	S
Rangayyan et al. (1997)	1997			√		54 r	S
Guliato et al. (1998, 2003)	1998			√		18 r	S
Kupinski and Giger (1998)	1998		√		√	118 r	S
Qi and Snyder (1998)	1998			√		11 r	S
Petrick et al. (1999, 2002)	1999		√	√		560(253) m	S
Lee et al. (2001)	2001			√		204 m	D and S
Chu et al. (2002)	2002			√		20 r	S
Kinnard et al. (2002, 2004b,a)	2002		√	√		124 m	S
Toledo Santos et al. (2002)	2002			√		109 r	S
Martí et al. (2003)	2003	√	√		√	20 r	S
Zhang et al. (2004) and Zhang and Foo (2006)	2004			√		40 m	D and S
Gulsrud et al. (2005)	2005		√	√		16 r	S
Hejazi and Ho (2005)	2005			√		60(80) m	D and S
Herredsvela et al. (2005)	2005		√	√		18 m	D
Sheshadri and Kandaswamy (2005)	2005		√	√			S
Wei et al. (2005, 2006)	2005		√	√	√	488(44) m	D and S
Eltonsy et al. (2007)	2007		√			270(135) c	D and S

Region growing algorithms have been widely used in mammographic mass segmentation with the aim to extract the potential lesion from its background. Since the early nineties, researchers from the University of Chicago investigated the introduction of shape information into the homogeneity criterion. With the aim to integrate the radiologists experiences, Huo et al. (1995) developed a semi-automatic region growing approach, in which the growing step was automatically computed after a radiologist had manually placed the seed point. Later, Kupinski and Giger (1998) compared this initial approach to two improved versions. The first one incorporated the Radial Gradient Index, which is a measure of the average proportion of gradient which are radially directed outwards (for a circular region the Radial Gradient Index is equal to one). The second one was based on a probabilistic method in which the probability of belonging to a region was modelled by a non-Gaussian distribution (using a kernel distribution), while the background was modelled using a uniform probability. They showed that the latter version improved performance compared to the other two approaches. Kinnard et al. (2002, 2004b,a) proposed a similar approach, where the cost function depended on the contours of the image. Subsequently, the boundary of the mass was located by using those points where a significant change in the cost function was found. In a similar probabilistic context, Martí et al. (2003) used active regions to model the mass in ROIs images. They modelled the mass as a Gaussian function, using features related to contour, texture, and shape information, and associated the background with a uniform probability.

Guliato et al. (1998) implemented a fuzzy version of the region growing algorithm. In contrast to the above approaches, no prior shape information was considered. Guliato et al.'s method was based on considering the uncertainty present around the boundaries of a tumour region, with the aim to preserve the transition between mass and normal tissue. An alternative approach was proposed by Bårman and Granlund (1994) who included edge information in the region growing algorithm in a multi-resolution framework. Petrick et al. (1999, 2002) also introduced gradient information into the region growing algorithm, but with the objective to reduce merging between adjacent and overlapping structures. Initially, the algorithm selected seeds using local maxima in the grey-scale image. In a subsequent step, a gradient image was constructed by using frequency-weighted Gaussian filtering. With this image, the thresholds of the regions bounded by the edges are extracted. Similar to the Chicago approach, Petrick et al. aggregate groups of pixels with similar characteristics (using thresholds), but, in contrast, they do not use shape information in the homogeneity criterion. However, this information was recently incorporated by Wei et al. (2005, 2006) who instead of using the frequency-weighted Gaussian filtering used an adaptive ring filter. Instead of segmenting the mass, Pohlman et al. (1996) segment the background region near the lesion and the boundary of the lesion is assigned to the inner boundary of this surrounding region.

Other researchers concentrated their efforts on improving the region growing algorithm by identifying the optimal set of initial seeds (mass detection). Zhang et al. (2004) and Zhang and Foo (2006) divided the breast in small regions and the pixels with maximum grey value were taken as seed points. Qi and Snyder (1998) used Bézier splines to interpolate histograms, from which they extracted threshold values at local maxima. Zheng et al. (1995b) used as a starting point the information coming from an edge image. This image was obtained by subtracting two Gaussian blurred images with a large difference in kernel sizes. Subsequently, they improved this work using a false positive reduction scheme (Zheng et al., 1999a) and including shape condition factors to limit the growing step (Zheng et al., 2003b). Another way to locate the initial starting points of the algorithm is to find the local maxima in the mammograms, which can be done using mathematical morphological operations, as suggested by Hejazi and Ho (2005), or small window thresholding, as proposed by Woods and Bowyer (1994).

A different approach was proposed by Chu et al. (2002), where the region growing approach is represented as a growing tree whose root is the selected seed. Active leaves are deleted in the connection area between adjacent regions to avoid merging adjacent structures. The authors stated that this graph-based segmentation more closely matches radiologists' outlines of masses.

2.1.2. Watershed methods

The watershed transform is a mathematical morphological approach to image segmentation (Beucher and Lenteuejoul, 1979; Vincent and Soille, 1991). Its name stems from the manner in which the algorithm segments the image regions into catchment basins (the low points in intensity). If water falls into these basins, the level in each basin rises until it is shared with its neighbouring basin. Thus, the output of this algorithm is a hierarchy of basins. What remains is to select the most discriminative level of basins for each purpose.

In mammographic mass detection and/or segmentation this algorithm has been used by different researchers. All of them noted that a pre-processing stage was necessary in order to reduce the number of initial segmented basins. Hence, Toledo Santos et al. (2002) used histogram enhancement, Herredsvela et al. (2005) and Gulsrud et al. (2005) applied image blurring, and Sheshadri and Kandaswamy (2005) suggested to extract the internal curvilinear structures (blood vessels, veins, milk ducts, speculations and fibrous tissue) before applying the watershed transform.

2.1.3. Split and merge methods

The split and merge technique (Chen and Pavlidis, 1979) is the other classical region-based segmentation method. As the name indicates, the process consists of recursively splitting the image until all regions satisfy a homogeneity criterion. In an accompanying step, all adjacent regions satisfying a second homogeneity criterion are merged. However, to our knowledge, for mammographic mass segmentation, this approach has only been used by Rangayyan et al. (1997) who, beginning with a hand-selected region of interest containing a single mass, used this technique to approximate its boundary using polygons.

2.2. Contour-based methods

Image segmentation techniques based on edge detection have been in use since the early work of Roberts (1965). However, identifying regions on the basis of edge information is far from trivial, since algorithms for edge detection do not usually possess the ability of the human vision system to complete interrupted edges using experience and contextual information. Therefore, sometimes edges are detected which are not the transition from one region to another and correctly detected edges often present gaps at places where the transitions between regions are not abrupt enough. Hence, detected edges may not necessarily form a set of closed connected curves that surround distinct regions.

As indicated in Table 2, there are only limited publications trying to detect masses using edge-based methods, which is mainly due to the difficulty of extracting the boundary between masses and normal tissue. Typical algorithms for finding edges are based on filtering the image in order to enhance relevant edges prior to the detection stage. The earliest approaches for mass detection are based on such methodology. The location of edges in Petrick et al. (1995, 1996b,a) was based on a Gaussian–Laplacian edge detector, after which the image was enhanced by an Adaptive Density-Weighted Contrast Enhancement Filter. Parr et al. (1994) used Gabor filters to locate the spicules of stellated lesions. The Hough transform (Davies, 1997) is a common approach in industrial

Table 2
Contour-based mammographic detection and/or segmentation techniques indicating used feature information and the purpose of each work. The penultimate column indicates the number of images used for validating the approach (r stands for ROI, m for mammogram, p for pair (2 mammograms), and c for case (4 mammograms)) and the parenthesis stand for the number of images used for training the approaches. The last column indicates the purpose of the described approaches: detection of masses in full mammograms (D), the segmentation of a mass given a small patch of the mammogram or a given seed point (S), or the detection and segmentation of the mass given the full image (D and S).

Author	Year	Texture	Gradient	Grey-level	Shape	Num testing	D/S
Parr et al. (1994)	1994		√	√		50 r	S
Petrick et al. (1995, 1996b,a)	1995		√	√		84(84) m	D and S
Groshong and Kegelmeyer (1996)	1996		√			44 m	D
Kobatake and Yoshinaga (1996)	1996		√	√		34(19) r	S
Zhang et al. (1996)	1996	√	√			42 m	D
Kobatake et al. (1998, 1999)	1999		√	√		1212 r	S
te Brake et al. (1999)	1999		√	√		132 m	S
Sahiner et al. (2001b,a)	2001		√	√		249 r	S
Allen et al. (2002)	2002		√	√		905 r	S
Hong and Brady (2003)	2003		√			48 m	D and S
Yin et al. (2003)	2003		√			20 m	D and S
Fauci et al. (2004, 2005) and Cascio et al. (2006)	2004		√	√		3762 m	D and S
Nakagawa et al. (2004)	2004		√	√		53 r	S
Timp and Karssemeijer (2004)	2004		√	√		3974 r	S
Shi et al. (2008)	2008		√	√		909 r	S

applications for object detection. For mammographic purposes, it had been used by Groshong and Kegelmeyer (1996) to detect circumscribed lesions, and also by Zhang et al. (1996), which used it to model tissue texture changes near spiculated lesions. A different approach is described by Kobatake and Yoshinaga (1996), which starting with a sub-image containing a possible mass lesion, looked for spicules using gradient information in three steps: firstly, the morphological line-skeletons were extracted in order to detect long and thin anatomical structures (like spicules). Secondly, a modified Hough transform was used to extract lines passing near the centre of the mass, and finally the algorithm automatically selected candidates based on the number of line-skeletons that satisfy the second step.

There have been recent attempts to avoid the necessity of filtering the image. In this sense, a topographic representation of the mammogram was proposed by Hong and Brady (2003) in order to detect salient regions. By thresholding at different topographic levels they were able to find different types of regions, like the pectoral muscle, breast density, or masses. A similar approach was used by Yin et al. (2003), who investigated the use of an intelligent mesh for finding the masses. The mesh is represented by a set of nodes and springs connecting them. The nodes are adapted to the edges of the image, and suspicious regions are those with a high density of nodes. Fauci et al. (2004, 2005) and Cascio et al. (2006) looked for the contours of the mass using an iterative algorithm. At each local maxima a threshold was selected which was used to draw an iso-intensity contour. The threshold value was based on user interaction and histogram information. Subsequently, the area of the selected region was refined by adjusting the threshold.

In other approaches, edge information has been used to refine initial segmentation results. Examples are Kobatake et al. (1999) and Sahiner et al. (2001b,a), who used active contour models (snakes) as a final step of their algorithms. Nakagawa et al. (2004) used two images for calculating the forces of the snakes, one related to edge intensity and the other based on grey-level information. Allen et al. (2002) and more recently Yuan et al. (2007) and Shi et al. (2008) used level sets for accurately finding the border of the lesions. Timp and Karssemeijer (2004) found the best contour of the mass by an optimisation technique based on dynamic programming. Their approach used both edge-based information as well as a priori knowledge about the grey-level distribution of the region of interest around the mass. They demonstrated a better performance of their method in comparison with an implemented version of the region growing algorithm inspired on the explained work of Kupinski and Giger (1998), and the discrete contour model inspired by the initial work of Lobregt and Viergever (1995) and adapted for mammographic purposes by te Brake et al. (1999).

There are other related approaches which are based on the detection of spicules and the use of statistical analysis of gradient-orientation maps (Karssemeijer and te Brake, 1996; te Brake and Karssemeijer, 1999; Zwiggelaar et al., 1999a). However, due to the necessity to perform a posterior classification step we consider these as model-based approaches.

2.3. Clustering methods

Clustering methods are one of the most commonly used techniques for image segmentation, as discussed in the review by Jain et al. (1999), and also for mass detection and/or segmentation, as can be seen from the reviewed approaches shown in Table 3. Based on the work of Jain et al., clustering techniques can be divided into hierarchical and partitional algorithms, where the main difference between them is that *hierarchical methods* produce a nested series of partitions while *partitional methods* produce only a single partition. Although hierarchical methods can be more accurate, partitional methods are used in applications involving large datasets, like the ones related to images, as the use of nested partitions is computationally prohibitive. However, partitional algorithms have two main disadvantages: (1) the algorithm has to know, a priori, the number of regions that are in the image, and (2) clustering algorithms do not use spatial information inherent to the image.

The traditional partitional clustering algorithm is the *k*-Means algorithm (MacQueen, 1967), which is characterised by simple implementation and low complexity. For mass segmentation, this algorithm has been used by Sahiner et al. (1996, 1998) in order to generate initial segmentation results. As we have described in the previous section, Sahiner et al. improved this segmentation using edge information. In contrast, Li et al. (1997, 1999, 2002) used a generalisation of *k*-Means that included spatial information (based on the work of Pappas (1992)) to refine an initial detection (which is achieved by using adaptive thresholding in a wavelet multi-resolution framework). Qian et al. (1998a,b, 1999, 2001) investigated how the pre-processing steps affect the result of the segmentation algorithm used by Li et al. and its posterior classification performance.

The Fuzzy *C*-Means (FCM) algorithm (Bezdek, 1981) is an extension of the *k*-Means algorithm which allows each pattern of the image to be associated with every cluster using a fuzzy membership

Table 3
Clustering based mammographic detection and/or segmentation techniques indicating used feature information and the purpose of each work. The penultimate column indicates the number of images used for validating the approach (r stands for ROI, m for mammogram, p for pair (2 mammograms), and c for case (4 mammograms)) and the parenthesis stand for the number of images used for training the approaches. The last column indicates the purpose of the described approaches: detection of masses in full mammograms (D), the segmentation of a mass given a small patch of the mammogram or a given seed point (S), or the detection and segmentation of the mass given the full image (D and S).

Author	Year	Texture	Gradient	Grey-level	Shape	Num testing	D/S
Brzakovic et al. (1990)	1990			√		25(10) m	D and S
Kobatake et al. (1994) and Kobatake and Murakami (1996)	1994			√		15 m	D
Gupta and Undrill (1995) and Undrill et al. (1996)	1995	√	√	√		31 r	S
Li et al. (1995)	1995	√		√		95(20) m	D and S
Comer et al. (1996)	1996	√		√		36 m	D and S
Laine et al. (1996)	1996	√		√			S
Miller and Ramsey (1996)	1996			√		67 c	D and S
Sahiner et al. (1996, 1998)	1996			√		168 r	S
Sameti and Ward (1996) and Sameti et al. (1997)	1996			√		35 m	S
Chen and Lee (1997)	1997	√				41 m	S
Li et al. (1997, 1999, 2002)	1997	√		√		34(60) m	D and S
Matsubara et al. (1997, 1998)	1997			√		231 m	D and S
Polakowski et al. (1997)	1997		√	√		254(18) m	D and S
Goto et al. (1998)	1998		√	√		18 r	S
Qian et al. (1998a,b, 1999, 2001)	1998	√				100 r	S
Giménez et al. (1999)	1999			√			D
Pfisterer and Aghdasi (1999, 2001)	1999	√		√		56 m	S
Rogova et al. (1999)	1999	√		√			S
Zheng et al. (1999b) and Zheng and Chan (2001)	1999	√				322 m	D
Heath and Bowyer (2000)	2000	√				160(156) m	D
Rocha et al. (2000)	2000		√	√		25 r	S
Velthuizen (2000)	2000			√		199 r	S
Mudigonda et al. (2001)	2001		√	√		56 m	D and S
Xie and Ma (2001) and Xie (2002)	2001	√		√			S
Kasai et al. (2002)	2002	√		√		626 m	D
Khan et al. (2002)	2002	√		√		30 m	D
Kwok et al. (2002)	2002	√				28 m	S
Ball et al. (2004)	2004			√		62 r	S
Cao et al. (2004a,b)	2004			√		123 r	S
Catarious et al. (2004)	2004			√		183 r	S
Abdel-Dayem and El-Sakka (2005)	2005			√		48 m	D and S
Özekes et al. (2005)	2005			√		52 r	D and S
Bruynooghe (2006)	2006			√		95(5) m	D and S
Kom et al. (2007)	2007			√		61 m	D and S
Varela et al. (2007)	2007			√		66 (30) m	D and S
Rojas and Nandi (2008)	2008			√		57 m	D
Suliga et al. (2008)	2008			√		100 m	D and S

function. FCM was used with different objectives in the works of Velthuizen (2000) and Chen and Lee (1997). Velthuizen used it to group pixels with similar grey-level values in the original images, while Chen and Lee used it over the set of local features extracted from the application of a multi-resolution wavelet transform and Markov Random Fields (MRF) analysis (Bishop, 2006). Moreover, the output of the FCM was the input of an Expectation Maximisation (EM) algorithm (Dempster et al., 1977) based on Gibbs Random Fields (Bishop, 2006). These final steps are closely related to the algorithm proposed by Comer et al. (1996).

In contrast to FCM which improves *k*-Means using a fuzzy approach of the energy function, the Dogs and Rabbit (DaR) algorithm (McKenzie and Alder, 1994) performs a more robust seed placement. The DaR was used by Zheng et al. (1999b) and Zheng and Chan (2001) to obtain an initial set of regions which subsequently were used to initialise a MRF approach. As Li et al. stated (Li et al., 1995), MRF allow the modelling of joint distributions in terms of local spatial interactions, introducing thus, local region information into the algorithm. This information was also introduced in the work of Rogova et al. (1999) using a constrained stochastic relaxation algorithm with a disparity measure function, which estimated the similarity between two blocks of pixels in the feature space. In contrast, Cao et al. (2004a,b) used two information theory based clustering algorithms to segment masses. The first approach was the Deterministic Annealing approach (Rose, 1998), which is a global minimisation algorithm and incorporated "randomness" into the to be minimised energy function. In the second approach, they unified a fuzzy based clustering and Deterministic Annealing to obtain an improved algorithm. In contrast with these approaches, Bruynooghe (2006) segmented an enhanced image instead of the original mammogram. The enhanced image was obtained by removing the locally linear fine detail structure using a morphological algorithm based on successive geodesic openings (Davies, 1997) with linear structuring elements at various orientations.

One of the earliest approaches to mass detection was the work of Brzakovic et al. (1990), which was based on a multi-resolution fuzzy pyramid linking approach, a data structure in which the input image formed the basis of the pyramid and each subsequent level (of lower resolution) was sequentially constructed. The links between each node and its four parents were propagated using a fuzzy function to upper levels. They demonstrated that this algorithm was directly correlated with the isodata clustering algorithm (Brzakovic et al., 1990). It has to be noted, that with this strategy, spatial information (region information) is taken into account.

Like Fu and Mui (1981), we consider *threshold methods* as partitional clustering methods. Threshold methods have been widely used for mass detection and/or segmentation. For instance, a threshold can be applied to obtain an initial rough detection of suspicious regions and, in a subsequent step, the result is refined using, for example topological analysis (Giménez et al., 1999) or the already explained region growing and snakes algorithms.

Most of the thresholding algorithms are based on the grey-level value of the actual mammogram. For instance, Abdel-Dayem and

El-Sakka (2005) found the best threshold to detect masses based on minimising the global fuzzy entropy of the image. In contrast, Matsubara et al. (1997, 1998) used different threshold values depending on the type of tissue of the breast, which was analysed using histogram analysis. Goto et al. (1998) refined the result of this work with the aim to detect possible spicules of the mass, and Kasai et al. (2002) included a pre-processing stage for detecting masses near the skin-line.

Mudigonda et al. (2001) used multilevel thresholding to detect closed edges. In this approach a concentric group of contours represents the propagation of density information from the central-core portion of an object or tissue region in the image into the surrounding tissues. This algorithm can be regarded as a region growing approach, where in each iteration neighbours with similar grey-level values are grouped (the works of Huo et al. (1995), as well as Petrick et al. (1999), described in subsection 2.1, follow a similar strategy). The main drawback of this approach is the assumption that masses have (more or less) uniform density compared to the local background. On the other hand, Özekes et al. (2005) used directional thresholding to estimate the shape of the mass. The pixels of the mammogram were scanned in eight directions using various thresholds. Subsequently, a (circular) mass template was used to categorise the region as being a true mass. Sameti and Ward (1996) and Sameti et al. (1997) introduced neighbouring information into the thresholding algorithm. They first divided the image into a set of ROIs and, subsequently, a fuzzy membership was given to each pixel of the ROI. In each iteration an error value was calculated, updating also each membership value. In this process they took neighbouring values into account.

In other cases the thresholding is not applied directly to the mammographic image, but to an enhanced version of the original image. For example, Ball et al. (2004) threshold a contrast enhanced version of the mammogram. Kobatake et al. (1994) and Kobatake and Murakami (1996) applied an Iris filter designed to enhance rounded opacities and to be insensitive to thin anatomical structures. Using adaptive thresholding they detected round masses. A similar approach has been recently used by Varela et al. (2007). In the work of Khan et al. (2002), the threshold to segment the data of an inner disk was computed using the background grey-level located inside an outer ring. Xie and Ma (2001) and Xie (2002) automatically located the threshold to segment a probability density function coming from wavelet analysis and Bayesian estimation of the mammogram. A logical filter was used by Rocha et al. (2000) in order to enhance the edges of the suspicious region. Subsequently, a thresholding value was found using histogram information to find the edge of the lesion.

Instead of enhancing the image using filters, a different approach is to first extract some features from the image and threshold them in a posterior step. For instance, Heath and Bowyer (2000) developed a new mass detection algorithm which was based on an Average Fraction Under the Minimum (AFUM) filter, which was designed to find the degree to which the surrounding region of a point radially decreases in intensity. The final step was to threshold the image to identify suspicious regions. Another example is the work of Gupta and Undrill (1995) and Undrill et al. (1996) who thresholded the mammograms using Laws masks. Pfisterer and Aghdasi (1999, 2001) showed that this last approach can be improved by combining it with wavelet decomposition. This work is similar to that proposed by Laine et al. (1996), where the best scale to segment the lesions was also found using a similar wavelet decomposition (Laine et al., 1995). In contrast to using a wavelet decomposition, Miller and Ramsey (1996) proposed a non-linear multi-scale analysis by maximum entropy. The aim of this analysis was to eliminate unwanted confusing structures.

Another common approach is to threshold the result of an image subtraction. Polakowski et al. (1997) found the edges in the image by subtracting two smoothed versions of the original mammogram, and subsequently thresholded this image. Two Gaussian filters with different standard deviation were used to smooth the original mammogram. In recent work, Catarious et al. (2006) investigated how the parameters of this algorithm affect the true and false positive rates and the description of the detected suspicious regions. Catarious et al. (2004) also proposed the use of an iterative procedure to refine the result of this subtraction. Instead of subtracting two blurred images of the mammogram, Kom et al. (2007) subtracted from the original image a linearly transformed enhanced image, and Kwok et al. (2002) the reconstructed image coming from the Daugman–Downing texture demodulation.

2.4. Model-based methods

As model-based segmentation we define those methods that include a training stage to *learn* the specific objects to be detected. Subsequently, the system has to be able to detect and classify new images depending on the presence or absence of similar objects. The training step cover examples with and without the object present, where in our case, the object is a mass. Thus:

- From mammograms containing a mass, the system learns the probable location and the variation in shape and size of the mass.
- From mammograms not containing a mass, the systems can learn features that represent normality.

Based on both training aspects the system learns what features to look for when presented with a new image. Table 4 shows publications based on such a strategy.

One of the most commonly used model-based segmentation methods is pattern matching. In pattern matching, the training is usually based on images containing the object to detect. Since the early proposal of Lai et al. (1989), pattern matching has been used for mass detection in mammographic images by different researchers (Ng and Bischof, 1992; Che et al., 1996; Constantinidis et al., 1999, 2000, 2001) using the normalised cross-correlation distance as the similarity measure. The main drawback of these approaches is the difficulty to account for the large variation in the shapes of masses. A different similarity measure that can be used to determine if a query ROI contains a true mass is mutual information, as used by Tourassi et al. (2003). In contrast, Oliver et al. (2006) and Freixenet et al. (2008) proposed to use a probabilistic template matching scheme to detect masses. The shape and deformations of a deformable template were learnt from real mass examples. Subsequently, a Bayesian scheme was used to adapt the learnt deformable template to the real contours of the mammogram. On the other hand, Hatanaka et al. (2001) successfully used a similar approach to detect masses with a partial loss of region, i.e. those masses located at the border of the image or on the boundary between the pectoral muscle and the breast.

Chang et al. (1996b,a) established a simple approach for detecting suspicious regions based on five rules. The selected regions should contain: (1) a global maximum in a Gaussian smoothed image; (2) a local maximum in the original image; (3) a local maximum in the image coming from the subtraction of two smoothed images (one using a Gaussian filter and the other using a box filter); and either (4) a small suspicious region of low contrast; or (5) a small suspicious region of high contrast. This approach was later combined with the initial approach of Zheng et al. (1995b) using logical operations (Zheng et al., 1996): the *or* operation increased the sensitivity of both algorithms as well as the number of false positives, while the *and* operation drastically decreased

Table 4
Model-based mammographic detection and/or segmentation techniques indicating used feature information and the purpose of each work. The penultimate column indicates the number of images used for validating the approach (r stands for ROI, m for mammogram, p for pair (2 mammograms), and c for case (4 mammograms)) and the parenthesis stand for the number of images used for training the approaches. The last column indicates the purpose of the described approaches: detection of masses in full mammograms (D), the segmentation of a mass given a small patch of the mammogram or a given seed point (S), or the detection and segmentation of the mass given the full image (D and S).

Author	Year	Texture	Gradient	Grey-level	Shape	Num testing	D/S
Lai et al. (1989)	1989			√	√	17 m	D
Kegelmeyer (1992) and Kegelmeyer et al. (1994)	1992	√	√			85 c	D and S
Ng and Bischof (1992)	1992			√	√	27 m	D
Karssemeijer (1994, 1999) and Karssemeijer and te Brake (1996)	1994		√	√		50 m	D
Stathaki and Constantinides (1994)	1994	√					S
Tarassenko et al. (1995)	1995	√	√	√	√	16(24) m	D and S
Calder et al. (1996)	1996		√				S
Chang et al. (1996b,a)	1996			√		510 m	D
Che et al. (1996)	1996			√	√	40 m	D
Diahi et al. (1996)	1996			√			D
Li et al. (1996b,a, 2001)	1996	√		√		200 m	D and S
Kalman et al. (1997)	1997	√		√		55(13) m	D
Jiang et al. (1998)	1998		√		√	24 r	S
te Brake and Karssemeijer (1998, 1999)	1999		√	√		132 m	D
Zwiggelaar et al. (1998, 1999a)	1998		√	√		56 m	D and S
Constantinidis et al. (1999, 2000, 2001)	1999			√	√	470(180) m	D
Morrison and Linnett (1999)	1999		√			2 m	D
Christoyianni et al. (2000)	2000	√		√		74(44) m	D
Hatanaka et al. (2001)	2001		√	√		335 m	D
Liu et al. (2001)	2001		√	√		38(19) m	D
Lo et al. (2002)	2002	√		√		200 (124) m	D and S
Youssry et al. (2003)	2003	√		√		22 m	D
Campanini et al. (2004)	2004			√		512 m (1400 r)	D
Cheng and Cui (2004)	2004	√				100 m	D
Hassanien et al. (2004)	2004			√		4 m	D
Öktem and Jouny (2004) and Ali and Hassanien (2006)	2004	√			√	60(180) m	D and S
Mousa et al. (2005)	2005			√		322 m	D
Oliver et al. (2006) and Freixenet et al. (2008)	2006		√		√	120 m (912 r)	D
Sakellaropoulos et al. (2006)	2006			√		90(166) m	D
Székely et al. (2006)	2006	√		√		160 (100) m	D and S

the false positive rate, but also the sensitivity was less than using both algorithms independently.

There is a series of approaches which model the masses using statistical approaches. For instance, Zwiggelaar et al. (1999a, 1998) detected spiculated lesions as the union of two techniques: the first one modelled the centre of the mass using a directional recursive median filter, while the second technique modelled the surrounding pattern of linear structures applying a multi-scale directional line detector. The combination of both methods results in a probability image and the detection is performed by thresholding the resulting probability image. Li et al. (1996b,a, 2001) first applied an image enhancement algorithm using morphological filtering. Subsequently, they employed a finite generalised Gaussian mixture (FGGM) distribution to model the histogram. They incorporated the EM algorithm to determine the optimal number of image regions and the kernel shape in the FGGM model. The final step was the use of Bayesian relaxation labelling to perform the selection of suspected masses. This work was extended (Lo et al., 2002) in order to extract the boundary of spiculated masses. In contrast with Tarassenko et al. (1995) used Parzen windows (Duda et al., 2001) because, as they stated, this required fewer assumptions compared to Gaussian mixtures. Calder et al. (1996) compared the performance of Spatial Planar Models and Gibbs Field models for both micro-calcification and mass segmentation. They found that the latter approach outperformed the former at the cost of introducing a priori knowledge information. In similar work, Morrison and Linnett (1999) compared the performance of parametrising both micro-calcifications and masses using a Gaussian function or a Hyperbolic Secant (*sech*) function. They found that the latter approximation outperformed the Gaussian approach, which suffers from a large number of false positives due to the fact that noise spikes are more likely to be accepted with this model than using the *sech* function.

A common approach in model-based mass segmentation methods is to extract gradient information from the mammogram and subsequently use this information for training a classifier. Due to the training aspect we do not classify such approaches as pure edge-based methods. One of the early approaches using this strategy was the work of Karssemeijer (1994, 1999) and Karssemeijer and te Brake (1996). They first detected spicules using second-order Gaussian derivatives operators. The specific orientation of spicules and background pixels were estimated. With this information they constructed two new features that formed the input for the classification stage. In more recent work (te Brake and Karssemeijer, 1999, 1998), they improved this algorithm using a multi-scale approach. Another approach using edge information is by Jiang et al. (1998), which is based on the enhancement of the spicules using morphological operations and, subsequently, two features representing the concentration of spicules were used to train a classifier based on a discrimination function.

In a similar way, not only using gradient but also including texture features in modelling the suspicious regions, Kegelmeyer (1992) and Kegelmeyer et al. (1994) trained a binary decision tree, obtaining as a final output a probability image. In a recent approach, Székely et al. (2006) also used a decision tree to classify a sliding window to contain mass or normal tissue. Subsequently, a Markov Random Field is used to refine the obtained segmentation. In similar work, Liu et al. (2001) decomposed the image using wavelets and at each resolution, extracted a set of features, including an edge orientation histogram. Subsequently, from coarse to finer resolutions, each pixel was classified by using a binary decision tree. As a natural extension of the work of Aylward et al. (1998) who used Gaussian Mixture Modelling (GMM) (Duda et al., 2001) to segment the breast according to their tissue, Sakellaropoulos et al. (2006) used wavelet analysis and feature extraction to classify the pixels of the dense region as mass or normal tissue.

In contrast with all these approaches, Campanini et al. (2004) proposed not to extract specific features to detect the lesions, but instead, they trained a Support Vector Machine (SVM) using grey-level ROI information from mass and normal tissue samples.

Finally, there is a large set of approaches based on Neural Network classifiers. These usually formulated the problem of segmentation as a classification of ROI as suspicious or not. The features for training are intensity or texture related information based on a set of known ROI containing masses and a set of random samples from normal tissue. For instance, Stathaki and Constantinides (1994) trained a neural net classifier using features extracted from a two-dimensional autoregressive model. Wei et al. (1994) used a convolution neural network, which is a back-propagation neural net that directly operates on images, instead of the typical feature-based back-propagation neural nets, like the one proposed by Diahi et al. (1996). Kalman et al. (1997) proposed the use of a linear output sequential recursive auto-associative memory where hidden units are connected in feedback to the context input–output units in order to predict the next preceding feature. Christoyianni et al. (2000) used a radial-based function neural network to classify features derived from the histograms of each ROI. Öktem and Jouny (2004) combine a back-propagation neural network and a self-organising map trained using fractal analysis and spatial moments distributions. Hassanien et al. (2004) and Ali and Hassanien (2006) tested a pulse-coupled neural network, which has the ability to extract edges, image segments, and texture information from images. Cheng and Cui (2004), Youssry et al. (2003), and Mousa et al. (2005) all used fuzzy neural networks.

3. Mass detection using multiple views

The comparison of different mammographic images of the same person is common practice in breast radiology. Usually, this comparison is between

- left and right mammograms,
- two mammographic views (CC and MLO) of the same breast,
- same view mammograms taken at different times.

Table 5 shows the approaches found in the literature arranged according to the used comparison. Note that we do not classify the algorithms according to detection or segmentation purposes because the result of all these algorithms is the detection of suspicious regions, not the perfect segmentation of the boundary of the lesions.

This practice is justified from a clinical point of view by Kopans (1998), who makes two important observations when comparing different mammograms of the same woman:

- though one breast may be larger than the other, internal structures are quite symmetric over broad areas, and

Table 5
Mammographic mass detection techniques using more than one image, detailing the used approach and the number of images used for validation (*r* stands for ROI, m for mammogram, *p* for pair (2 mammograms), and *c* for case (4 mammograms)). The parenthesis stand for the number of images used for training the approaches.

Author	Year	Left–right	CC-MLO	Temporal	Num testing
Lau and Bischof (1991)	1991	√			114 m
Yin et al. (1991, 1993, 1994) and Giger et al. (1992)	1991	√			46 c
Brzakovic et al. (1994)	1994			√	32 p
Sallam and Bowyer (1994, 1996)	1994			√	1 p
Zheng et al. (1995a)	1995	√			500 m
Stamatakis et al. (1996)	1996	√			40 p
Vujovic et al. (1996) and Vujovic and Brzakovic (1997)	1996			√	29 p
Zouras et al. (1996)	1996			√	79 c
Highnam et al. (1998) and Kita et al. (1998, 2001)	1998		√		32 p
Karssemeijer and te Brake (1998)	1998	√			132 p
Kok-Wiles et al. (1998)	1998	√			26 p
Méndez et al. (1998, 2003)	1998	√			70 p
Chang et al. (1999), Good et al. (1999)	1999		√		571 p
Good et al. (2001)	1999			√	
Sanjay-Gopal et al. (1999) and Hadjiiski et al. (2001b,a)	1999			√	74 p
Bovis and Singh (2000)	2000	√			144 m
Leung and Sickles (2000)	2000	√			84615 m
Martı et al. (2001, 2002)	2001			√	3 p
Sun et al. (2001, 2004)	2001		√		20(30) p
Attikiouzel and Chandrasekhar (2002)	2002	√			
Paquerault et al. (2002)	2002		√		169 p
Wirth et al. (2002)	2002			√	
Georgsson (2003)	2003	√			100 p
Richard and Cohen (2003)	2003	√			3 p
Van Engeland et al. (2003)	2003			√	150 p
Zheng et al. (2003a)	2003			√	260 p
Christoyianni et al. (2004)	2004	√			20 p
Altrichter et al. (2005)	2005		√		188 c
Filev et al. (2005)	2005			√	318 p
Timp et al. (2005) and Timp and Karssemeijer (2006)	2005			√	389 p
Wai and Brady (2005)	2005			√	
van Engeland et al. (2006), van Engeland and Karssemeijer (2006, 2007)	2006		√		412 c
Wang et al. (2006)	2006	√			3 p
Hachama et al. (2006)	2006	√			3p
Martí et al. (2006)	2006	√			64 p
Zheng et al. (2006)	2006		√		450 c
Qian et al. (2007)	2007		√		100 p
Pu et al. (2008)	2008		√		200 p
Velikova et al. (2009)	2009		√		268 p

- overlapping tissue structures that form summation shadows and normal tissue variations on the mammogram highlight unimportant asymmetries.

In order to distinguish masses and asymmetric breast tissue, clinicians take a range of characteristics of abnormal areas, such as size, density, and shape into account.

The following subsections provide a more detailed description of the various approaches to automatic mass detection from multiple mammographic images.

3.1. Comparison of left and right mammograms

The comparison between left and right breasts is based on the fact that both breasts have a similar internal structure. Fig. 2 shows mammograms corresponding to the left and right breast of a woman. A tried and tested method is known as *bilateral subtraction*. Both images are first aligned and subsequently subtracted. The alignment of both breasts is the critical component in this process. There are two main approaches to this:

- The use of anatomical features, like the position of the nipple, internal regions or assumptions on the compression of the breasts (Yin et al., 1991, 1994; Méndez et al., 1998, 2003; Bovis and Singh, 2000; Attikiouzel and Chandrasekhar, 2002; Georgsson, 2003; Richard and Cohen, 2003; Christoyianni et al., 2004; Wang et al., 2006).
- The segmentation of the profile of both breasts (Giger et al., 1992; Karssemeijer and te Brake, 1998).

The drawback of the latter techniques is that these are not likely to register the breast interior correctly. The main reason is that these techniques do not take the distortion of the internal structures into account. Note that compression conditions are likely to be different for the left and right mammograms. In the literature there are studies comparing the performance of bilateral subtraction with single-image algorithms (Yin et al., 1993; Zheng et al., 1995a). While the former study suggested that bilateral subtraction performs better than single-image detection, the conclusion of the latter contradicts this. Both techniques are non-exclusive and can be used in parallel to improve the results (Karssemeijer and te Brake, 1998).

There are alternatives to this alignment and subtraction approach. Kok-Wiles et al. (1998) represent the bright zones of the breast as a tree-structure, which forms the basis for the comparison process. Stamatakis et al. (1996) developed a *Multiple Image Comparison* approach. This method starts with one pair of mammograms, subsequently the system computes eight new pairs of images using a set of grey-level features. Finally each pair is bilaterally compared and the resulting images are recombined into a final result.

3.2. Comparison of two mammographic views

The comparison of two views of the same breast is also known as *ipsilateral comparison*. Fig. 3 shows two different views of the same breast. The common approach is to firstly detect masses in both mammograms independently using a single view approach, and afterwards, map the suspicious regions on the other view by using a correspondence algorithm. Therefore, a comparison is made in order to ensure that the suspicious detected regions are real masses or false positives.

Differences between the approaches can be found in both steps of the algorithms. For instance, the first step of these algorithms can be based on each of the reviewed methods in Section 2. On the other hand, the correspondence algorithm can be used in different ways: (1) to validate that a suspicious region detected in both views is the same (Paquerault et al., 2002; Pu et al., 2008), (2) to help an algorithm to detect a mass in the second view once a possible mass have been detected in the first view (for instance, by reducing the decision threshold of a probability image) (Zheng et al., 2006), or (3) to extract joint features of both views for a posterior classification step (Sun et al., 2004; van Engeland et al., 2006).

The correspondence algorithms rely on the fact that radiologists use the distance to the nipple to correlate a lesion in both MLO and CC views (van Engeland et al., 2006; Wei et al., 2009) since it is generally believed that this distance remains fairly constant. Hence, once the nipple is (automatically) located a polar coordinate system is centred on its position (if the nipple does not appear in the image the furthest point to the chest wall can be used instead Zheng et al., 2006). To validate if a mass is the same in both views the distance from the mass to the nipple is computed and compared (Paquerault et al., 2002; Pu et al., 2008). In contrast, other algorithms open a search strip where to look for the possible masses (Sun et al., 2004; van Engeland et al., 2006; Zheng et al., 2006).

Two recent approaches are the works of Pu et al. (2008) and Velikova et al. (2009). The former used the skin-line for fitting an ellipse based model whose major and minor axis defined the directions of a Cartesian coordinate system located at the point where one of the axis intersects with the breast boundary. This way, instead of using a polar coordinate system, this approach allows the use of a Cartesian coordinate system. On the other hand,

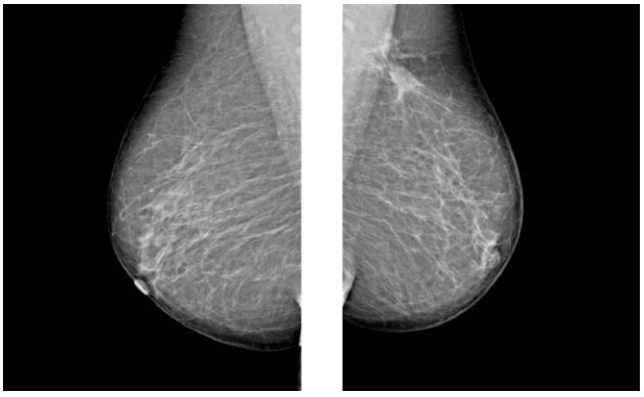

Fig. 2. Right and left mammograms of a woman (selected from the Trueta database).

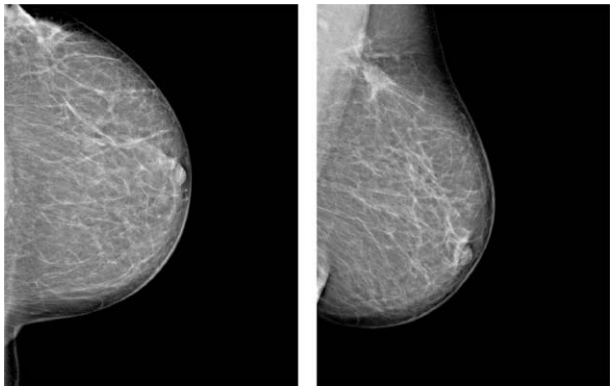

Fig. 3. CC (left) and MLO (right) views of the same breast (selected from the Trueta database).

Velikova et al. (2009) proposed a Bayesian network framework that takes into account not only the lesions region, but also the links between regions.

3.3. Temporal comparison of mammograms

Finally, a comparison can be performed between mammograms of the same breast but temporally separated. Radiologists use the same approach to evaluate how a suspicious region has evolved. Sanjay-Gopal et al. (1999) first identify regions of interest on the most recent mammogram. Subsequently, the nipple location is used to align mammograms and locate the detected region in the previous mammogram. Vujovic and Brzakovic (1997) proposed an algorithm for identifying the potential control points (for instance crossing points) and establishing the correspondence between them. Using a similar methodology, Martí et al. (2001) extracted salient control points from internal linear structures in order to establish the correspondence. The latter provides a more robust approach to the extraction of control points using features such as orientation and width of the structures, instead of just their positions. Recently, Wai and Brady (2005) proposed to extract the landmarks at the junctions of curvilinear structures, including those generated from connective tissue, vessels and ducts. See Van Engeland et al. (2003) for a quantitative comparison of some of the mammographic registration methods.

4. Quantitative comparison of mass detection methods

In Section 2, mass detection and/or segmentation approaches using one single view have been qualitatively analysed and classified as region, contour, clustering or model-based methods. However, based only on this analysis, we cannot estimate the robustness of the algorithms with respect to different mammographic cases (different breast tissue, different lesion types and sizes). Therefore, in order to study how the performance varies for the different strategies, and with the aim to extract reliable conclusions, we have quantitatively compared seven different approaches, which are shown in Table 6. These approaches are selected as being representative of the main state-of-the-art general purposes detection techniques, but at the same time, taking into account the fact that they have been successfully applied to our particular problem, which is the detection (and some of them segmentation) of masses, and showing good performance. Moreover, we tried to represent the proportion of algorithms for each category: one region-based and one contour-based (which are the least used strategies), two model-based (which is the intermediate used strategy), and three based on clustering (the most used strategy). The performance of the different approaches has been analysed using both ROC analysis and FROC curves (Metz, 1996) to evaluate the detection/segmentation performance. A comparison between all the algorithms is provided to determine their main advantages and constraints.

4.1. Data

The algorithms have been evaluated using two different databases: a publicly available digitised database of mammograms (the MIAS database) and a full-field digital mammograms database (the Trueta database). Subsequent subsections describes both databases in more detail.

For the mammograms of the MIAS database a pre-processing step has been performed in order to remove the background and labels. The breast and pectoral muscle area were segmented from the background using an adaptive threshold (Raba et al., 2005). This segmentation results in a minor loss of skin-line pixels in the breast area, but those pixels are not relevant for mass detection. Note that this step is not necessary for the mammograms belonging to the Trueta database since this database contains full-field digital mammograms and hence, the background is homogeneous and there do not appear any labels in the images. Subsequently, a second pre-processing step has been performed in order to identify the pectoral muscle. Clearly, this approach limits the detection/segmentation of masses to the breast area, but this was used as some of the implemented algorithms use or recommend to use such approach (Karssemeijer and te Brake, 1996; Petrick et al., 1996b). In order to detect the pectoral muscle region, a seed is automatically located in this area and using an adaptive region growing algorithm it is grown until the pectoral muscle is detected. The main advantage of using these two pre-processing approaches is the ease of implementation and fast performance.

4.1.1. MIAS database

The first mammographic database used is a subset of 261 mammograms extracted from the Mammographic Image Analysis Society (MIAS) database (Suckling et al., 1994). This is a publicly available digitised dataset with the following characteristics: resolution of 50×50 μm, grey-scale response linear in the optical density range 0–3.2 OD, and images quantised to 8 bits. For decreasing the computational cost, all images were reduced by a factor of 4 using Gaussian downsampling twice. Of this set, 54 mammograms contained masses, whilst the other 207 represented normal cases (see Table 7 for a breakdown of the dataset). The 54 abnormal contained all the MIAS images annotated as spiculated, circumscribed or miscellaneous (ill-defined masses). The following were not included: *mdb059ls* because the abnormality location is not included in the database and *mdb290rs* because the abnormality is located on the boundary of the mammogram. In addition, *mdb069ll* which

Table 7
Summary of the MIAS mammograms used in this work.

	B-I	B-II	B-III	B-IV	Total
Circumscribed	9	6	3	2	20
Spiculated	4	7	8	1	20
Ill-Defined	7	4	3	0	14
Normal	56	67	58	26	207
Total	76	84	72	29	261

Table 6
Compared mass detection methods, detailed according to the used strategy. DoG is referred to Difference of Gaussians.

Strategy	Id	Method	Based on	D/S
Region	*a*1	Region growing	Eltonsy et al. (2007)	D and S
Contour	*b*1	Laplacian	Petrick et al. (1996b)	D and S
Clustering	*c*1	Thresholding	Kom et al. (2007)	D and S
	*c*2	Iris filter	Kobatake and Murakami (1996)	D
	*c*3	DoG	Polakowski et al. (1997)	D
Model	*d*1	Pattern matching	Lai et al. (1989)	D
	*d*2	Classifier	Karssemeijer and te Brake (1996)	D and S

was annotated as circumscribed was considered as spiculated based on the detailed annotations by the expert radiologist. With the aim to simulate a more realistic scenario, all mammograms labelled as normal were included in the dataset. All circumscribed and spiculated lesions have been manually segmented by a breast screening radiologist and the annotations are used as ground-truth data. The ground-truth for the ill-defined masses is located using the location and the radius provided in the MIAS annotations. Moreover, all mammograms have been classified according to BIRADS density categories by three expert radiologists with the final classification found by majority voting.

4.1.2. Trueta database

The second database originated from the Radiologic Department of University Hospital Dr. Josep Trueta. The database contains cases, where each case can be composed of both MLO and CC views. Each image of the database (acquired using a Siemens Mammonat Novation) has 70 μm pixels quantised to 12 bits. Two different image sizes depending on the breast size are included: 2560×3328 or 3328×4096 pixels. As in the MIAS database, all mammograms were reduced by a factor of 4 for computational reasons. Each case has its own annotations, including the breast density classified using BIRADS categories and if a lesion is present, its boundary (accurately outlined by an expert radiologist).

Table 8 shows the distribution according to the BIRADS breast density categories. The database contains 89 MLO and 87 CC views containing masses, and 70 MLO and 74 CC normal mammograms.

4.2. Implemented mass detection methods

As shown in Table 6, each selected method is symbolised using the first letter of the class and a number to distinguish between each algorithm. This description will be useful when presenting results and comparing algorithms. The following subsections provide a detailed description of the selected algorithms.

4.2.1. a1: Based on a detection of concentric layers

Eltonsy et al. (2007) proposed the detection and segmentation of masses by locating the presence of concentric layers with progressively lower average intensity. The algorithm begins using a region granulation: a grey-level transformation to reduce the large number of intensity levels. Thus, pixels strongly connected in terms of both spatial location and grey-level intensity are grouped together in the same granule level.

The granulation step begins by linearly normalising the intensity between 0 and 1. Following, the pixels are assigned to a granule level. This is done by sequentially visiting each pixel and examining its local, 3×3, neighbourhood. If all neighbours are within 98% of the granularity, they are assigned to the same granule level, otherwise they will be assigned to a different level. After this granulation step, a morphological opening is performed to reduce scattered granule levels.

Table 8
Summary of the Trueta mammograms used in this work.

	B-I	B-II	B-III	B-IV	Total
MLO					
Masses	46	21	20	2	89
Normal	34	11	23	2	70
Total	80	32	43	4	159
CC					
Masses	45	20	20	2	87
Normal	36	14	22	2	74
Total	81	34	42	4	161

The detection and segmentation of suspicious regions is based on the inspection of all the granule levels. This way, for each level, all the regions with the same or higher level are grouped, and a set of features, including area, eccentricity, solidity, and dispersion are computed. This process is repeated for the brightest levels. The features are used to limit the number of feasible layers. Afterwards, the concentric layers are found, and the suspicious regions are those containing at least three evolving concentric layers. All the parameters used have been empirically adjusted to our databases. One of these parameters is related to the minimum distance between possible masses and this distance can be used to obtain probability images.

4.2.2. b1: Based on a Laplacian edge detector approach

The *b1* algorithm is inspired by Petrick et al. (1996b), Petrick et al. (1996a), who used an "optimal" Laplacian–Gaussian edge detector in order to find closed regions in an enhanced image of the mammogram. The algorithm starts by pre-processing the image using a Density-Weighted Contrast Enhancement (DWCE) filter. This filter is based on two filtered images of the original mammogram $I(x,y)$:

- A density image $F_D(x,y)$, which is an smoothed version of the mammogram, and is found by applying a Gaussian filter.
- A contrast image $F_C(x,y)$, which is found by subtracting the original image from a second smoothed version of this image.

The density image is filtered again using a non-linear filter K_M and used to define a multiplication factor which modifies the corresponding pixel in the contrast image. This way, it allows the local density value of each pixel to be weighted by local contrast. This intermediate image F_{KC} can be analytically described as:

$$F_{KC}(x,y) = K_M(F_D(x,y)) * F_C(x,y) \quad (1)$$

This image is used to define a second multiplication value using another non-linear filter K_{NL}, which is multiplied again by the weighted contrast of the corresponding pixels:

$$F_E(x,y) = K_{NL}(F_{KC}(x,y)) * F_{KC}(x,y) \quad (2)$$

The resulting image $F_E(x,y)$ is the output of the DWCE filter. The result of this filtering process is an image with the potential masses highlighted.

To select the masses an edge detector is applied. A Laplacian–Gaussian edge detector was used, which defines edges as the zero crossing locations of:

$$\nabla^2 G(x,y) * F_E(x,y) \quad (3)$$

where $G(x,y)$ is a two-dimensional Gaussian smoothing function with standard deviation σ, which is the only parameter for the detection. The detector is used recursively from low resolution (large σ) to high resolution (small σ) using the edge map of the previous stage as a guide for the current edge detection.

Once the edge detector has been applied a set of regions are found, some of these are suspicious regions while the remaining regions represent normal tissue. In order to distinguish both types of regions, a set of simple morphological features are computed. The thresholding of these features determines which regions are retained and this information forms the basis for the ROC and FROC analysis.

4.2.3. c1: Based on thresholding

This algorithm is inspired on the work of Kom et al. (2007). The method is based on thresholding an image obtained by subtracting from the mammogram a linear filtered representation of itself. Firstly an enhanced image $E(x,y)$ is found by using:

$$\begin{cases} E(x,y) = a\log(1+mI(x,y)) & I(x,y) < \alpha \\ E(x,y) = (\exp(I(x,y)/a) - 1)/b & I(x,y) > \alpha \end{cases} \quad (4)$$

where $I(x,y)$ is the original mammogram, m is its maximum grey-level, b depends on the values of a and m: $b = (1 - \exp(m/a))/m$, and a and α are two parameters found experimentally (in the original work, 10,000 and 0.3, respectively). The idea of this filter is to enhance both dark and light areas). Following, the subtracted image $S(x,y)$ is found by:

$$S(x,y) = I(x,y) - E(x,y) \quad (5)$$

Finally, the image S is thresholded by using an adaptive local threshold to obtain the suspicious regions. Variation in the threshold values results in different number of regions and this forms the basis for the ROC and FROC analysis.

4.2.4. c2: Based on Iris filter

This approach is based on enhancing rounded opacities by means of the analysis of a gradient-orientation map (Kobatake et al., 1994; Kobatake and Murakami, 1996), which is obtained by using the so-called Iris filter. As shown in the recent work of Varela et al. (2007) the output of this filter strongly depends on the area considered to calculate it.

The Iris filter is applied to the gradient image obtained as the maximisation of the first-order directional derivative (we used the 3×3 Prewitt operators). For each pixel P in the gradient image, N half lines radiating from it are considered. For every half line, points within a distance R_{min} and R_{max} are considered. For each of these points (Q_i) the convergence degree of the gradient vector at that point is defined as:

$$C_i(x,y) = \frac{\sum_{Q=P_{min}}^{Q_i} f(Q)}{\overline{P_{min}Q_i}}, \quad Q_i \in [P_{min}, P_{max}] \quad (6)$$

where P_{min} and P_{max} are the first and last points of the line being considered, and the function $f(Q)$ is defined for each pixel as:

$$f(Q) = \begin{cases} \cos(\theta) & |g| \neq 0 \\ 0 & |g| = 0 \end{cases} \quad (7)$$

where θ is the orientation of the gradient vector at the point with respect to the half line being considered. The maximum convergence degree of the half line is $C_{io} = \max(C_i)$. Hence, the output of the Iris filter at pixel P is defined as the average of C_{io} on the N half lines:

$$C(P) = \frac{1}{N}\sum_{i=0}^{N-1} C_{io} \quad (8)$$

The result of this filter is that rounded opacities, such as masses, are highlighted, while linear structures are mostly removed. Moreover, the use of the minimum radius allows the algorithm to avoid enhancing micro-calcifications.

The result of this step is a pseudo-probability image in the range of $[-1,1]$, where higher values are obtained near the centre of round convex regions. Hence, thresholding these images at different levels are used for the ROC and FROC analysis.

4.2.5. c3: Based on a Difference of Gaussians

The approach of Polakowski et al. (1997) combined thresholding and edge features. The first step consists in highlighting the possible masses by subtracting two smoothed versions of the original mammogram. Subsequently, this image is thresholded to locate the masses. Catarious et al. (2006) also investigated how the parameters of this algorithm modifies the true and false positive rates and the description of the detected suspicious regions.

The core of this algorithm is the use of the Difference of Gaussians (DoG) filter. It is constructed by subtracting two Gaussians with different standard deviations:

$$\mathrm{DoG}(r) = \frac{1}{2\pi\sigma_1^2}\exp\left\{\frac{-r^2}{2\sigma_1^2}\right\} - \frac{1}{2\pi\sigma_2^2}\exp\left\{\frac{-r^2}{2\sigma_2^2}\right\} \quad (9)$$

This filter acts like a bandpass filter, and frequencies inside both standard deviations are highlighted. Hence, masses of different sizes can be found by varying these values. After the filtering process, a global threshold is applied in order to detect the masses. The optimisation of this threshold is database dependent. Note that a posterior false positive reduction step is necessary in order to discard normal tissue. The values in the subsequent ROC and FROC analysis are obtained by varying this threshold.

4.2.6. d1: Based on a pattern matching approach

Pattern matching starts by defining a template, in our case, a tumour-like template. The definition of the template is based on the approach of Lai et al. (1989), who defined the tumour by three characteristics: brightness contrast, uniform density and circular shape. In our implementation, the template can vary between 3 and 201 pixels in diameter, in increments of 3 pixels. Fig. 4 shows a 5-pixel radius template. The circular patch of ones in the centre represents a tumour area having uniform density. The ring of zeros represents the "do not care" area to account for some of the shape variability. Finally, the outer edge of the template is filled with minus ones to represent the dark background. One of the drawbacks of this algorithm is its poor performance in detecting spiculated masses (Lai et al., 1989).

The matching probability is calculated using mutual information (Tourassi et al., 2003). Given two images A and B, the mutual probability is expressed as:

$$MI(A,B) = \sum_{\alpha}\sum_{\beta} P_{AB}(\alpha,\beta)\cdot log_2\left(\frac{P_{AB}(\alpha,\beta)}{P_A(\alpha)P_B(\beta)}\right) \quad (10)$$

where $P_{AB}(\alpha,\beta)$ is the joint probability of the two images based on their corresponding pixels values and $P_A(\alpha)$ and $P_B(\beta)$ are the marginal probabilities of the variables α and β which are the image grey-level values at a pixel level. To obtain a compatible template we calculated the mean of all pixels in the breast area. Subsequently, in the template:

- -1s are replaced with pixels values inferiors to the mean: randomly in the range of 3/4–9/10 of the mean value,
- 0s with the value of the mean, and
- 1s with values superiors to it. The values of the template varies from the centre to the borders of the template. To be specific, the

```
        -1 -1 -1
     -1 -1  0 -1 -1
  -1 -1  0  1  0 -1 -1
-1 -1  0  1  1  1  0 -1 -1
-1  0  1  1  1  1  1  0 -1
-1 -1  0  1  1  1  0 -1 -1
  -1 -1  0  1  0 -1 -1
     -1 -1  0 -1 -1
        -1 -1 -1
```

Fig. 4. A tumour-like template for matching with tumours of five pixels in diameter (Lai et al., 1989). The ones in the centre represents a tumour area having uniform density, the zeros represents the "do not care" area to account for some of the shape variability, and the outer minus ones represent the dark background.

centre value is the grey-level value equal to 95/100 of the cumulative histogram of the full mammogram, while the values at the border are similar to the mammogram's mean grey-level value.

The result of this step is a probability image, with high values in those regions where suspicious structures are found. Hence, thresholding these images at different values provides the basis for the ROC and FROC analysis.

4.2.7. d2: Based on a classifier approach

The final approach uses classification and has been inspired by the work of Karssemeijer and te Brake (1996) and Karssemeijer (1999). The *d2* algorithm finds possible masses from the detection of spicules using second-order Gaussian derivatives operators. If a line-like structure is present at a given site, the method provides an estimation of the orientation of these structures, whereas in other cases the image noise will generate a random orientation. With this information two new features are constructed. The first feature represents the total number of pixels pointing towards the centre, while the second one estimates if these directions are circularly oriented. With these two features and a set of classified mammograms *d2* trains a binary decision tree. Subsequently, the decision tree can be used for mass detection in unseen mammograms.

The result of the decision tree is a likelihood image, where brighter regions means a higher degree of suspiciousness. Hence, thresholding these images at different levels the algorithm will obtain the results for the subsequent ROC and FROC analysis.

4.3. Evaluation methodology

The evaluation of CAD algorithms is performed by comparing the results obtained by the detection/segmentation algorithms with the expert(s) annotations, which are considered as the gold standard. Hence, each image is labelled as normal or abnormal (in our case as containing masses, independently of their diagnosis) by both the expert and the automatic algorithm. The manual annotations can be considered as binary images incorporating a detailed outline of the mass when present. In contrast, the result of the automatic algorithms is not a single binary image, but a probability image containing a certain number of grey-levels. In this kind of image, each grey-level represents the likelihood degree of being part of a mass. Hence, brighter pixels are those with high probabilities to be a mass, while darker pixels are those with high probabilities to belong to normal tissue.

Depending on the aim of the algorithm (detection/segmentation/diagnosis) different methodologies are used for the evaluation, as is indicated in Fig. 5. Our evaluation has been done in terms of ROC and FROC analysis. The starting point is a set of probability images resulting of applying one of the seven implemented algorithms to the dataset of mammograms. Thresholding all the probability images at the same threshold we obtain images with marked regions (mammograms with masses) and without marked regions (the normal mammograms). Comparing the resulting binary images with the set of manual annotations we obtain the measure commonly identified as "sensitivity at false positives per image". This measure consists of counting the overall number of correctly identified images with masses relative to all abnormal images (sensitivity), and counting the number of images detected as containing masses but actually being normal mammograms relative to all normal images (false positives per image). If at each threshold, the pair of sensitivity and false positives rate is plotted as a point in a graph (with the sensitivity on the *y*-axis and the false positives rate on the *x*-axis) we will obtain a Receiver Operating Characteristic (ROC) curve. A common measure derived from this curve is the area under the curve (A_z), which is an indication for the overall performance of the observer. For a perfect classifier

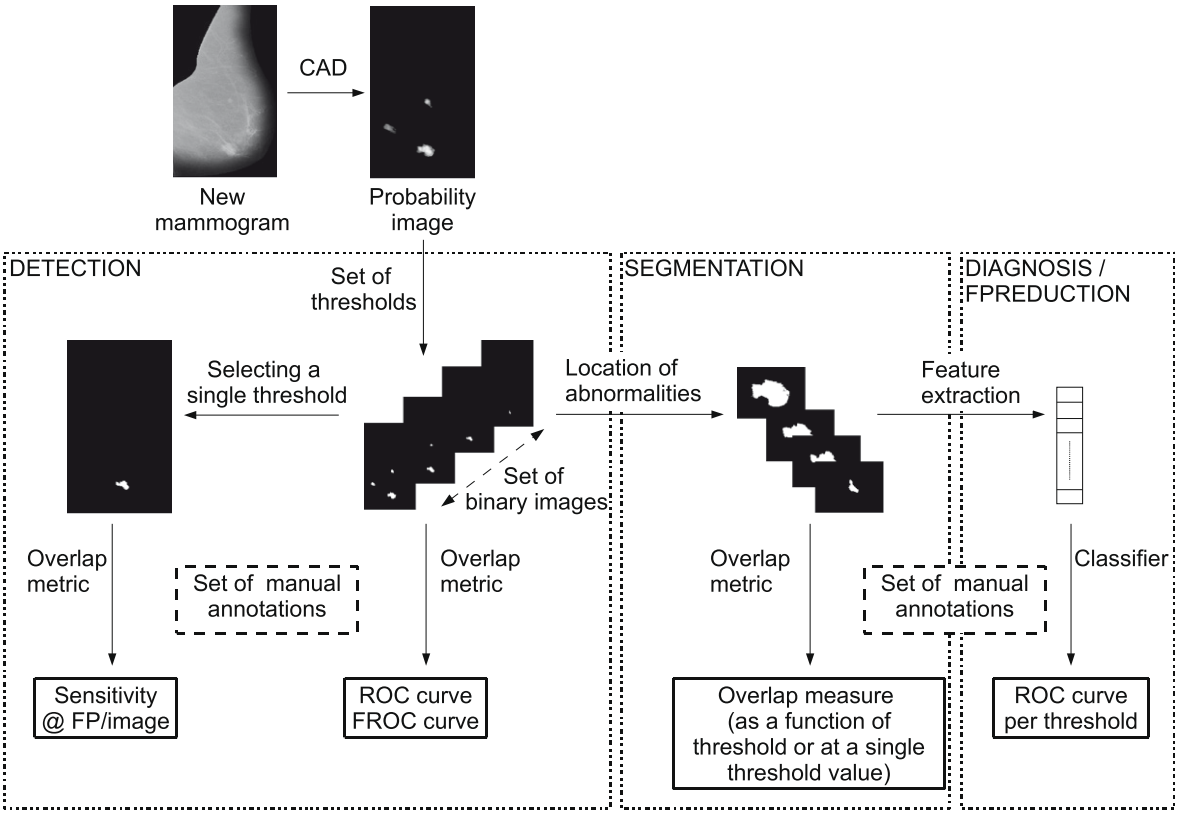

Fig. 5. The evaluation methodology depends on the aim of the CAD algorithm. Note that the result of the algorithm must be normalised for obtaining a proper comparison when thresholding the mammograms of the database.

the A_z value is equal to one. Note that increasing the number and spread of the threshold values in general results in a more accurate estimation of A_z. In this work we used 100 threshold levels to obtain the results.

However, note that using ROC analysis one algorithm can correctly classify an image as abnormal despite an incorrect localisation of the abnormality. In order to take this information into account, a region based analysis is performed using Free-response Receiver Operating Characteristic (FROC) curves (Metz, 1996). In this analysis the Lesion Localisation Fraction (LLF) is obtained as the number of correctly detected lesions relative to the total number of lesions and the Non-Lesion Localisation Fraction as the number of non-correctly detected lesions relative to the total number of images. The FROC curve is the graphical summary of both measures (Yoon et al., 2007). Note that the definition of a detected region is needed. Different approaches coexist in the literature, for instance, a mass is detected if the centre of gravity of the automatic annotation is inside the manual annotation (Székely et al., 2006; Zheng et al., 2003a), if the pixel with the highest measure of suspiciousness (probability) falls inside the manual annotation (Karssemeijer and te Brake, 1996; te Brake and Karssemeijer, 1999), or using an overlap criterion (Lai et al., 1989; Kegelmeyer et al., 1994; Petrick et al., 2002; Bornefalk, 2005; Freixenet et al., 2008).

For plotting the FROC curve, we look for the maximum sensitivity of each algorithm, and subdivide the resulting range to give 20 sensitivities and determine the number of false positives at these 20 sensitivities to obtain the points on the curve. In our evaluation we consider a correctly localised lesion if there is a 50% overlap between the automatic detection and the manual annotation. Although this definition seems clear it has some problems due to the fact that the algorithms return probability images and with the decrease of threshold values new regions appear at local maxima and are merged at local minima (or saddle points). Fig. 6 schematically depicts this aspect. Image (a) represents the result of a manually segmented mass, with a 1 representing the mass and the zeros representing the normal tissue. Image (b) represents the result of an automatic algorithm, where higher values represent those regions more likely to be a mass. Lets start the FROC analysis thresholding the image at a value of 9, as it is shown in (c). In terms of this analysis, there is one false positive in the image. In (d) the threshold is set to 8. The two marks depict the same false positive and a new correctly detected lesion, since the area overlap between the annotated and the automatically generated mark is 100%. We follow the analysis reducing again the threshold, as it is shown in (e), where the threshold is set to 7. Note that in this case the area overlap between the correctly detected lesion and the annotated mark reduces to 20%. According to the 50% definition this would be considered as a incorrectly detected region and two false positives would be returned. The extreme case is shown in (f), where the region containing the abnormal region is merged with a false positive region and the original 50% rule would return a single false positive and hence reducing the number of false positive per image. However, here we have segmented the original probability images before thresholding and the obtained regions are preserved in the FROC analysis. The segmentation is based on extracting convex regions (peaks) in the probability images (Zwiggelaar et al., 1999b), where small regions (less than 10 pixels using a 4-connected-neighbourhood) were not considered. Using a similar approach (ignoring the 10 pixels threshold) the image in Fig. 6b would be divided into two regions, each containing a local maximum (i.e. the 8 and 9 values). Using this approach, the results for Fig. 6e and f would be a single true positive and false positive.

The results in this paper were obtained using a 10-fold cross-validation methodology. Each dataset was divided in 10 different groups, nine of them were merged for training the algorithms and obtaining the best parameters for such group, while the remaining group was used to test the algorithms using the obtained parameters. For each fold this resulted in an A_z value for the training and test data. This procedure was repeated until all groups were used for testing. Finally, the ROC and FROC analysis were performed merging all the tested images. Note, that using this methodology, each mammogram appears in the test set only once, while it appears in the training set nine times (in order to not bias the results, note that for the Trueta database each training group must include the full patient case). This procedure was repeated three times in order to obtain an indication of the variation in the results (the resulting small standard deviations indicate that repeating the experiments three times is sufficient). The presented average and standard deviation results are based on the combination of the three times 10-fold cross-validation results (e.g. the average A_z values are based on the 30 A_z results from each fold).

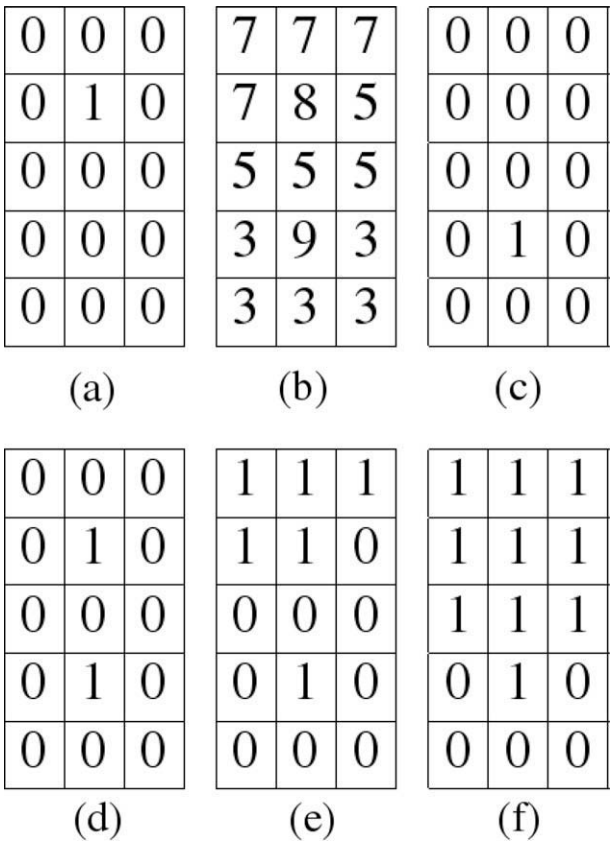

Fig. 6. Comparison of (a) a manually annotated image and (b) an automatically probability image resulting from an algorithm. (c)–(f) the results of segmenting the images at different thresholds (9,8,7,5).

4.4. Mass detection results

The initial experiment is related to the capability of the algorithms to distinguish mammograms with and without masses, i.e. the capability of the algorithms to detect masses. This aspect of the evaluation mimics the radiologist in identifying the presence of abnormalities. The ROC and FROC analysis and the cross-validation procedure as described above are used.

Table 9 shows the mean and standard deviation of the A_z values obtained when training the algorithms for both the MIAS and Trueta databases. In addition, the table shows the results obtained for the test data for both databases. As expected, the performance on training data shows better results than found for the test data. The difference in standard deviation between the MIAS and Trueta data could indicate a higher variation in the Trueta data.

Table 9
Results obtained by the algorithms using ROC analysis for both sets of mammograms. The results are given in terms of area under the curve (A_z). The last row shows the mean and standard deviation for all the algorithms.

	Train		Test	
	MIAS	Trueta	MIAS	Trueta
a1	0.630 ± 0.064	0.579 ± 0.077	0.614 ± 0.032	0.564 ± 0.061
b1	0.778 ± 0.014	0.729 ± 0.013	0.758 ± 0.005	0.686 ± 0.002
c1	0.710 ± 0.014	0.653 ± 0.013	0.685 ± 0.002	0.649 ± 0.005
c2	0.804 ± 0.015	0.760 ± 0.012	0.787 ± 0.003	0.757 ± 0.011
c3	0.674 ± 0.015	0.593 ± 0.022	0.601 ± 0.028	0.594 ± 0.043
d1	0.690 ± 0.017	0.755 ± 0.019	0.675 ± 0.005	0.715 ± 0.001
d2	0.708 ± 0.028	0.718 ± 0.012	0.673 ± 0.016	0.690 ± 0.004
M	0.713 ± 0.060	0.684 ± 0.076	0.685 ± 0.068	0.665 ± 0.068

Comparing the performance for both databases, note that for the training data all algorithms except *d1* and *d2* tend to decrease their performance when using the Trueta database. This trend is mirrored in the test data results. These differences can be a reflection on how the data has been obtained: digitised versus digital, the quality of the data, and the size and visibility distribution of the masses within the data. It should be noted that in the original publications, all algorithms were evaluated on digitised data (like the MIAS database).

On the other hand, when comparing the results on the test data for the same database, note that for MIAS, algorithms *b1* and *c2* obtained the best performance. Algorithm *b1* is based on the use of the DWCE filter to enhance lesions, while algorithm *c2* is based on the Iris filter. In contrast, for the Trueta database, the performance of the algorithms is more similar, although the Iris filter approach (*c2*) and pattern matching (algorithm *d1*) provide the best results.

The FROC results based on the set of 261 MIAS mammograms are shown in Fig. 7a. In general, all implemented approaches have a tendency to *over-detect* (to find more regions than the real ones) and hence to produce a large number of false positives at high sensitivity rates. Algorithms *b1*, *c2*, *d1* and *d2* show the best trade-off between sensitivity and false positives per image. Fig. 7b shows the FROC analysis based on the Trueta database. Note that again algorithms *c2* and *d2* show the best performance. However, all algorithms show a decreased performance when compared to the MIAS database. The performance of these two approaches is linked to the incorporation of directional distribution information. This aspect is expected to reduce the number of false positives which are likely to have a more random distribution than true positive regions. This indicates that spatial information is essential to reduce false positive regions to more acceptable levels.

Instead of using the full FROC curve, we can also compare the algorithms by establishing the best trade-off between sensitivity and false positives, i.e. looking for the best operating point of the algorithms. Note, however, that this is a subjective point since an expert can prefer to obtain less false positives (but poorer sensitivities) and another one can prefer higher sensitivities (at the expense of increasing the number of false positives). For instance, using the MIAS database, the operating point for algorithm *b1* can be established to a sensitivity equal to 0.790 at 4.064 false positives per image, or otherwise to a sensitivity 0.910 at 8.129 false positives per image. The same applies to algorithm *c2*, which corresponding operating point can be established to a sensitivity of 0.806 at 3.806 false positives per image or to 0.895 at 7.295 false positives per image. For algorithms *d1* and *d2* the best operating points can be established to sensitivity 0.754 at 6.044 false positives per image and sensitivity 0.823 at 6.353 false positive per image, respectively. On the other hand, testing the Trueta database, the operating points can be the following ones: for algorithm *b1* 0.704 at 7.585 false positives per image, for algorithm *c2* either

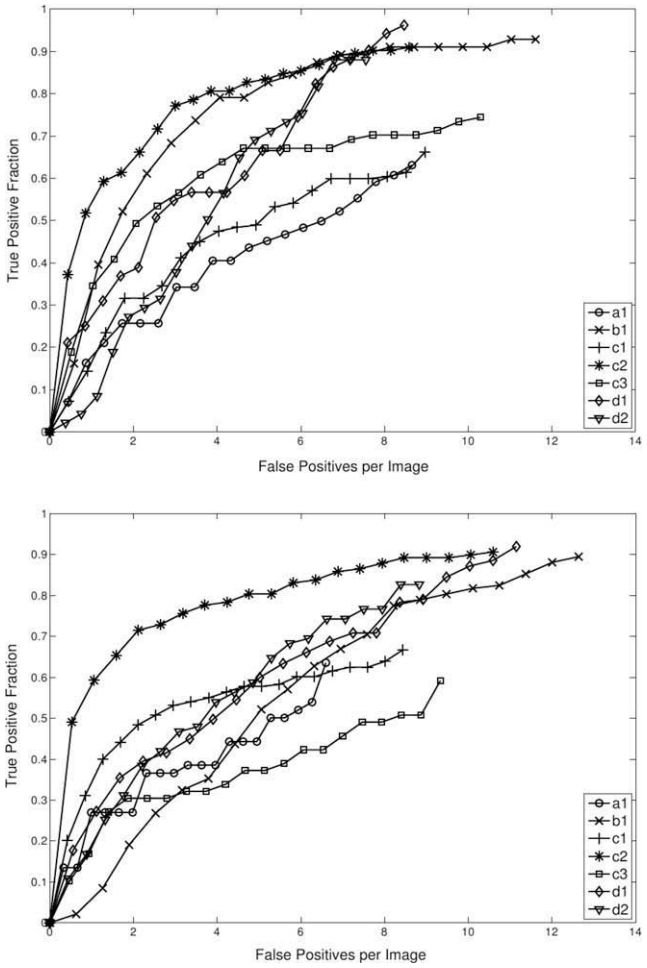

Fig. 7. FROC analysis of the algorithms using (a) the MIAS database and (b) the Trueta database.

0.753 at 2.894 false positives pe image or 0.855 at 6.271 false positives per image, for algorithm *d1* 0.728 at 6.007 false positive per image, and for algorithm *d2* 0.742 at 7.084 false positive per image.

Bornefalk (2005) and Bornefalk and Bornefalk-Hermansson (2005) presented a parametric methodology that allows estimation of the probability with which a particular CAD system performs better than another on unseen data in a clinical setting. It is based on modelling the sensitivity of an operating point as a binomial distribution while the false positives are modelled using a Poison distribution. Given a specific sensitivity, the methodology produces the 90% confidence interval for the number of false positives per image. For instance, using the MIAS database, at a sensitivity of 0.8, the number of false positives markings for the different algorithms is: *b1* (3.96,5.42), *c2* (3.11,5.17), *d1* (5.97,6.58) and *d2* (5.92,6.61) (the other algorithms did not achieve a sensitivity of 0.8 within the range of false positives per image that was investigated). These numbers clearly show that algorithms *b1* and *c2* produce significantly less false positives than the rest of algorithms. On the other hand, repeating this analysis for the Trueta database and the same sensitivity of 0.8 the false positives per image for the various algorithms are: *b1* (8.29,10.77), *c2* (3.29,5.49), *d1* (6.39,7.59) and *d2* (7.56,8.47). This indicates that using this digital database, algorithm *c2* significantly outperforms the other algorithms.

A qualitative comparison between these and the original results indicate the following. All algorithms showed a marked decrease when compared to the results presented in their original work.

Briefly, Eltonsy et al. (2007) (*a1*) indicated 92% sensitivity at 5.4 false positive per image testing 155 mammograms with masses and 82 without masses and using 164 training mammograms. Petrick et al. (1996b) (*b1*) detected 82 of 84 mammograms with masses, but obtained more than 10 false positives per image (this number was subsequently decreased by using a false positive reduction step). Kom et al. (2007) (*c1*) obtained $A_z = 0.95$ using 61 mammograms, 48 of them containing masses. Varela et al. (2007) (*c2*) obtained $A_z = 0.90$ at 3.4 false positive per image (again without the false positive reduction step) using 60 training mammograms and testing 130 mammograms with masses and 264 being normal ones. Polakowski et al. (1997) (*c3*) obtained 92% sensitivity at 8.38 false positives per image using 36 mammograms for the training and 236 for the testing. Lai et al. (1989) (*d1*) used 17 images containing abnormal regions (no normal mammograms were included in the study) and indicated a performance of 100% sensitivity at 1.7 false positives per image. Karssemeijer and te Brake (1996) (*d2*) used a combination of 9 stellated lesions, 10 architectural distortions and 31 normals mammograms to reach a sensitivity of 80% at about 1.0 false positive per image.

Comparing the reported performance with the obtained results it is clear that algorithms *a1*, *c1*, and *c3* drastically decrease their sensitivity, algorithms *c2*, *d1*, and *d2* maintain their sensitivity but increase the number of false positives, while the performance of algorithm *b1* is similar to the reported one. The poor performance of the former algorithms can be due to different factors, such the use of different datasets or the different normal to abnormal image ratio. For instance, note that algorithms *a1* and *c1* were evaluated with more mammograms containing masses than without containing masses. On the other hand, the increase in the false positive numbers in algorithms *c2*, *d1*, and *d2* was expected, since all these approaches used a posterior algorithm with the aim to reduce this number. Finally, note also that the number of false positive when using the Trueta database is larger than using the MIAS one. This is due to the fact that the digital technology allows a more detailed contrast of all the breast, and there appear structures highly contrasted that do not appear in digitised mammograms, which are associated with suspicious regions by the algorithms. Note, again, than these structures can be effectively eliminated using a false positive reduction strategy.

4.4.1. Breast tissue influence

The accuracy of the algorithms depending on breast tissue classification is summarised in Table 10 (the analysis is based on the test data). This table shows the mean A_z value for each algorithm detailed by each density. Hence, only mammograms corresponding to the same BIRADS category are used to obtain the values in the columns. The last row shows the mean and standard deviation for A_z for all the algorithms. In general, masses in low density breasts are better detected than masses in high density breasts, although each algorithm performs differently with regard to this.

For instance, looking at the MIAS results, algorithms *a1*, *c2*, and *c3* have superior performance on fatty breasts (BIRADS I) compared to the other density classes. In contrast, algorithms *c1* and *b1* obtained the best accuracy for mammograms belonging to BIRADS II and BIRADS III, respectively. Algorithms *d1* and *d2* performs better for the most dense mammograms (BIRADS IV). The Trueta based results show similar trends.

The reason for these differences in behaviour can be explained by different factors. For example, masses in fatty breasts usually have a more delineated boundary than in denser breasts, and also one can see a set of circumscribed layers around the mass that are exploited in the granularity approach (*a1*) and the Iris filter (*c2*). Moreover, as there are clear differences between the grey-level values of the mass and the rest of the breast, the Difference of Gaussians approach (*c3*) works well. Algorithms *b1* and *c1* have a filtering pre-processing step. This kind of approach seems beneficial for mammograms belonging to intermediate BIRADS (II and III) classes, where masses are highlighted with respect to the normal tissue. Finally, algorithms *d1* and *d2* performs better for the dense mammograms. Algorithm *d2* use contour information as a basis for the detection process and as such has a better performance when increased intensity changes are present. It should be noted that the results for the BIRADS IV class are based on a limited number of samples (see Tables 7 and 8).

4.4.2. Lesion shape influence

The lesion shape has a strong influence on the performance of the detection algorithms. Table 11 shows the sensitivity of the different algorithms detailed by the lesion shapes using the MIAS database. According to the annotations of this database, the lesions can be classified as Circumscribed (round and oval shapes), Spiculated (a mass with spicules), and Ill-defined (masses with inaccurate boundaries). There are, respectively, 20, 20, and 14 mammograms of each class (see Table 7).

As a global trend, algorithms show more accurate detection on circumscribed and spiculated masses than on ill-defined masses

Table 11
Influence of the lesion shape for the detection algorithms. The results show the sensitivity for mass detection detailing the behaviour according to their shape. The total number of masses of each class is: Circular: 20, Spiculated: 20, Ill-defined: 14. The last row shows the mean and standard deviation for all the algorithms.

	Circ.	Spic.	Misc.
a1	0.689 ± 0.017	0.625 ± 0.009	0.569 ± 0.007
b1	0.865 ± 0.002	0.762 ± 0.002	0.654 ± 0.002
c1	0.821 ± 0.002	0.632 ± 0.009	0.613 ± 0.001
c2	0.768 ± 0.007	0.809 ± 0.005	0.779 ± 0.008
c3	0.665 ± 0.009	0.636 ± 0.013	0.548 ± 0.015
d1	0.749 ± 0.001	0.692 ± 0.001	0.589 ± 0.001
d2	0.720 ± 0.008	0.648 ± 0.007	0.579 ± 0.007
Mean	0.754 ± 0.071	0.686 ± 0.072	0.619 ± 0.078

Table 10
Influence of the breast density for the detection algorithms. The results show the area under the curve (A_z) for each algorithm detailing the behaviour for the density categories classified according to the BIRADS standard. The last row shows the mean and standard deviation for all the algorithms.

	MIAS				Trueta			
	B-I	B-II	B-III	B-IV	B-I	B-II	B-III	B-IV
a1	0.671 ± 0.015	0.610 ± 0.017	0.590 ± 0.013	0.548 ± 0.010	0.652 ± 0.010	0.561 ± 0.012	0.560 ± 0.008	0.565 ± 0.007
b1	0.722 ± 0.005	0.743 ± 0.001	0.782 ± 0.001	0.695 ± 0.001	0.674 ± 0.002	0.651 ± 0.001	0.684 ± 0.001	0.689 ± 0.003
c1	0.658 ± 0.012	0.694 ± 0.001	0.667 ± 0.001	0.697 ± 0.007	0.642 ± 0.005	0.626 ± 0.002	0.681 ± 0.011	0.677 ± 0.016
c2	0.819 ± 0.013	0.815 ± 0.012	0.780 ± 0.014	0.680 ± 0.013	0.765 ± 0.006	0.742 ± 0.016	0.712 ± 0.005	0.681 ± 0.008
c3	0.613 ± 0.013	0.543 ± 0.054	0.611 ± 0.014	0.565 ± 0.011	0.641 ± 0.007	0.634 ± 0.014	0.556 ± 0.017	0.560 ± 0.017
d1	0.652 ± 0.006	0.653 ± 0.001	0.675 ± 0.001	0.823 ± 0.001	0.707 ± 0.002	0.716 ± 0.003	0.721 ± 0.002	0.806 ± 0.001
d2	0.646 ± 0.011	0.714 ± 0.008	0.639 ± 0.011	0.733 ± 0.009	0.627 ± 0.012	0.599 ± 0.013	0.602 ± 0.012	0.675 ± 0.006
Mean	0.683 ± 0.068	0.682 ± 0.089	0.678 ± 0.076	0.677 ± 0.095	0.673 ± 0.049	0.647 ± 0.063	0.645 ± 0.071	0.665 ± 0.084

Table 12
Influence of the lesion size for the detection algorithms using the Trueta database. The results show the sensitivity for mass detection detailing the behaviour according to their size. The number of masses of each class is: Ultra-Tiny (<1.0 cm^2): 72, Tiny (1.0—2.0 cm^2): 36, Small (2.0—3.0 cm^2): 12, Medium (3.0—4.0 cm^2): 13, Large (4.0—7.5 cm^2): 26, Extra-Large (>7.5 cm^2): 16. Last row shows the mean and standard deviation for all the algorithms.

	ut	t	s	m	l	xl
a1	0.572 ± 0.009	0.527 ± 0.008	0.582 ± 0.005	0.572 ± 0.009	0.562 ± 0.009	0.612 ± 0.004
b1	0.709 ± 0.003	0.687 ± 0.002	0.685 ± 0.002	0.720 ± 0.002	0.669 ± 0.002	0.654 ± 0.005
c1	0.626 ± 0.001	0.615 ± 0.005	0.626 ± 0.001	0.624 ± 0.002	0.673 ± 0.000	0.719 ± 0.002
c2	0.777 ± 0.003	0.782 ± 0.003	0.809 ± 0.004	0.788 ± 0.006	0.770 ± 0.004	0.764 ± 0.006
c3	0.610 ± 0.006	0.621 ± 0.009	0.638 ± 0.004	0.626 ± 0.004	0.615 ± 0.004	0.623 ± 0.005
d1	0.792 ± 0.003	0.715 ± 0.002	0.710 ± 0.002	0.742 ± 0.002	0.602 ± 0.004	0.581 ± 0.009
d2	0.782 ± 0.008	0.706 ± 0.005	0.674 ± 0.009	0.668 ± 0.008	0.608 ± 0.007	0.607 ± 0.002
Mean	0.695 ± 0.092	0.664 ± 0.084	0.675 ± 0.073	0.677 ± 0.076	0.643 ± 0.068	0.651 ± 0.067

(see last row of the table, which shows the mean of the algorithms). This is due to the fact that ill-defined masses have irregular and poorly defined borders, which tend to blend into the background. However, this is not the case for algorithm *c2*, which shows high A_z values for all types of masses. The algorithms *b1*, *d1*, and *d2* all perform well on the detection of circumscribed and spiculated masses.

4.4.3. Lesion size influence

The influence of lesion size on detection accuracy is summarised in Table 12. This table shows the sensitivity of the algorithms detailed by the lesion size, using the Trueta database (a similar behaviour is found using the MIAS database). According to the annotations and classifying the masses in the following sizes: *Ultra-Tiny* (<1.0 cm^2), *Tiny* (1.0–2.0 cm^2), *Small* (2.0–3.0 cm^2), *Medium* (3.0–4.0 cm^2), *Large* (4.0–7.5 cm^2), *Extra-Large* (>7.5 cm^2), there were 72, 36, 12, 13, 26, and 16 masses in each interval, respectively.

Observe that in the table there is not a specific size where all the algorithms do well nor a size where all the algorithms perform poorly, but algorithms vary their performance with lesion size. However, both algorithms *b1* and *c2* perform well regardless of the lesion size. Algorithms *d1* and *d2* perform better on the smaller lesions, while algorithm *c1* performs better on larger lesions. This limited performance can be related to the lack of gradient information associated with larger masses or to the variation in the training and test samples.

5. Discussions

We have presented and reviewed different approaches to the automatic and semi-automatic detection and segmentation of mammographic masses. Specific emphasis has been placed on the different strategies. A classification of both detection and segmentation techniques has been proposed, describing several algorithms and pointing out their specific features. Further, we have evaluated seven of the most frequently used strategies for mass detection using a single view. These methods have been fully evaluated using ROC and FROC analysis and tested using a commonly used digitised database and a new full-field digital database. The annotations, which were used as the gold standard, were provided by three expert mammographic radiologists.

It should be made clear that none of the investigated approaches provide the overall best detection performance. The analysis of the results using ROC analysis shows that the best performance, with $A_z = 0.762$ and $A_z = 0.780$, is obtained testing the MIAS database when using the Laplacian approach applied over an enhanced version of the mammogram or the thresholding approach after applying the Iris filter, respectively. In contrast, the best performances when testing the full-field digital mammogram are obtained again by the Iris filter approach, $A_z = 0.759$, and by the pattern matching approach, with $A_z = 0.715$. The obtained results show a dependence on the breast density. Testing the MIAS database, abnormal mammograms belonging to BIRADS I tend to show improved detection over abnormal mammograms belonging to other BIRADS category. This is related to the increase in parenchymal tissue, which is mistaken for abnormal regions. In contrast, when testing the Trueta database, this global trend is less clear, and the dependency varies for each algorithm. The lesion shape also affects the final results of the algorithms. As is shown in Table 11 the algorithms tend to find well-defined circumscribed masses easier to detect, followed by spiculated masses and finally ill-defined masses. Finally, we note that the results of the algorithms also depend on the lesion size, although no clear trend is evident.

The combination of algorithms is a powerful manner to obtain more robust systems for object detection. In this sense, we combined the results of the four algorithms with best performance using addition and multiplication of the corresponding probability images (applied to the MIAS database). Although there are more complex options to combine them, using these simple techniques indicated the possibilities for further development. The result when using addition was $A_z = 0.802 \pm 0.003$ and when using multiplication was $A_z = 0.800 \pm 0.004$. Note that for both combinations, the results outperform the detection performances when testing the algorithms individually (see Table 9). This shows that, effectively, the combination of algorithms can be a way to obtain improved sensitivities, and that this topic might be explored as future work.

On the other hand, the analysis of the FROC curves shows that algorithms based on the Laplacian, the Iris filter, the pattern matching, and the gradient analysis provide the best results, obtaining a sensitivity better than 0.8, although at this sensitivity the number for false positive per image is more than 5. However, this number could be decreased by using false positive reduction algorithms. These algorithms are based on the analysis of the suspicious ROIs in order to classify them as real masses or normal tissue. Usually, these algorithms are based on extracting a number of different features (based on grey-level or texture) and use them to distinguish between both types of ROIs (Sahiner et al., 1996; Qian et al., 2001; Lladó et al., 2009). Note, for instance, that in Varela et al. (2007), after the thresholding step, a false positive reduction algorithm based on extracting grey-level and morphologic features and training a neural net is proposed. A different approach to reduce the number of false positives is the use of information coming from another mammographic view (van Engeland and Karssemeijer, 2007; Wu et al., 2007).

We can compare the results obtained in this review with the results reported in the literature for commercial CADs. We should distinguish between evaluation per case and per image (as we do here). Sensitivities for the first group ranges from 0.70 to 0.90 at false positive rates between 0.35 and 2.6 FPs/case (Wu et al., 2007; Ellis et al., 2007; Yang et al., 2007; Wei et al., 2007), while for the second group the sensitivity is reduced to 0.43–0.80 at false

positive rates between 0.2 and 1.4 FPs/image (Ellis et al., 2007; Kim et al., 2008). These values tend to improve when micro-calcification detection is also included, due to the fact that the performance for commercial mass detection algorithms is lower than that of micro-calcification detection algorithms (Warren et al., 2000). Comparing these results with those obtained here we note that the sensitivity obtained is in the same range than the commercial ones. In contrast, the number of false positives at these sensitivities is larger. However, the use of the above mentioned false positive reduction algorithms should be useful for reducing this number. Note that commercial CAD systems include both false positive reduction step and potentially other processing stages to refine the initial detection results.

The superior performance of algorithms *b1* and *c2* when testing the MIAS database could be due to the fact that both algorithms apply pre-processing enhancement, showing that this kind of pre-processing algorithms could be beneficial. Note however, that changing the database the performance of both algorithms decreased (more drastically the Iris based approach). This should indicate that enhancing techniques are database dependent, which is an undesirable behaviour. A possible way to eliminate such dependency should be an initial analysis of the mammograms of the database, in order to automatically tune the possible enhancement parameters. Note that this is what model-based algorithms implicitly do. Moreover, digital technology allows us to easily obtain additional mammograms. However, the fact that ground-truth has to be manually provided by expert radiologists limits the use of such model-based algorithms. Although this is becoming less of a problem as more manually segmented mammograms can be found in the public domain, it should be noted that the quality of annotations is variable.

As shown in Section 4.4 lesion shape, size and tissue type strongly influence the performance of these algorithms. Few algorithms make use of breast tissue information which can be automatically known a priori before applying the detection algorithm (Byng et al., 1996; Karssemeijer, 1998; Zhou et al., 2001; Oliver et al., 2008). For instance, Matsubara et al. (1997) first classify a mammogram according to its density, and subsequently removing the dense cases in their mass detection approach. Eltonsy et al. (2007) adjust the training and testing databases according to the number of mammograms of each density class. In a recent proposal based on a pattern matching approach, Freixenet et al. (2008) train different templates for each density class. In the matching step, the mammogram is firstly classified in a density class, and only the corresponding templates are used for detecting masses. They report a similar performance for each class.

Our validation of the algorithms has been performed using both a digitised and a full-field digital database. Note however, that actual trends in mammography show that digital technology outperforms traditional analog imaging, in terms of quantum efficiency and higher resolution (Smith, 2003), acquisition time (Berns et al., 2006), low patient radiation (Obenauer et al., 2003), easier storage and transmission (the DICOM protocol, National Electrical Manufacturers Association, 2006) and also for the improved integration with computerised tools. We have used a (relatively) small set of digital mammograms. This is due to the fact that, up to date, there is not a publicly available full-field mammographic database, yet. Note that when one of these exists (and become an evaluation standard in mammography), future evaluation should shift from digitised to digital images.

6. Conclusions and future trends

Mass detection and segmentation techniques, in general, are still in need of improvement. In this sense, mammographic mass detection/segmentation algorithms using a single image are expected to improve in different ways. Actual trends in unsupervised object recognition are based on probabilistic latent semantic analysis and visual vocabularies (Sivic et al., 2005; Bosch et al., 2008). Although these approaches are outperforming the traditional supervised schemes for specific applications, we believe that supervised approaches are likely to achieve a better performance for mass detection/segmentation as such approaches are capable of including the large variation in the size and morphology of these abnormalities. Potentially successful detection/segmentation methods should combine both bottom-up information of the query image and top-down constraining information of the object to be detected (Kumar et al., 2005). Thus, a statistical model combining a priori information of masses, including grey-level, texture, and also shape information has to be combined with a detection/matching procedure including actual information of the query image (like, for instance, the breast tissue). In addition, boosting procedures (Freund and Schapire, 1997; Viola and Jones, 2001; Torralba et al., 2004) are expected to provide improved performance for methods including classifiers.

Finally, the integration of detection results from ipsilateral, bilateral and temporal mammograms is expected to bring improvements. The main idea in bilateral and temporal comparisons is to detect differences of tissue between the different views. Note that this information can be introduced in the algorithm a posterior, i.e. using it as a false positive reduction method, or also as a priori information. For example, it can be used as an enhancement strategy in order to highlight those regions which are different in both views and to remove similar regions. Moreover, in probabilistic mass detection algorithms, this information could be introduced as a penalising term in those regions where differences have been found. On the other hand, information coming from ipsilateral comparison seems more useful as a posterior information, i.e. to test if a suspicious region in one view has a corresponding match in the other view.

Although current results obtained by both commercial and research CAD systems make them a realistic tool in early mass detection, as indicated above there are still research questions related to enhancement, detection, segmentation, false positive reduction algorithms, and multiview integration which makes automatic mass detection and segmentation in mammography an active area within computer vision.

Acknowledgements

The authors would like to thank the reviewers for their critical comments that improved the content and readability of the paper considerably. This work was supported in part by Ministerio de Educación y Ciencia of Spain under Grant TIN2007–60553 and by the CIRIT and CUR of DIUiE of Generalitat de Catalunya under Grant 2008SALUT2009.

References

Abdel-Dayem, A.R., El-Sakka, M.R., 2005. Fuzzy entropy based detection of suspicious masses in digital mammogram images. In: IEEE Conference on Engineering in Medicine and Biology Society, pp. 4017–4022.

Ali, J.M.H., Hassanien, A.E., 2006. PCNN for detection of masses in digital mammogram. Neural Network World 16 (2), 129–141.

Allen, B.H., Oxely, M.E., Collins, M.J.A., 2002. A universal segmentation platform for computer-aided detection. In: International Workshop on Digital Mammography, pp. 164–168.

Altrichter, M., Ludányi, Z., Horváth, G., 2005. Joint analysis of multiple mammographic views in CAD systems for breast cancer detection. In: Lecturer Notes in Computer Science, vol. 3540, pp. 760–769.

American Cancer Society, 2007. Breast Cancer: Facts and Figures 2007–08, ACS.

Astley, S.M., Zwiggelaar, R., Parr, T.C., Taylor, C.J., 1998. Prompting in mammography: how good must prompt generators be? In: International Workshop on Digital Mammography, pp. 347–354.

Attikiouzel, Y., Chandrasekhar, R., 2002. DSP in mammography. In: International Conference on Digital Signal Processing, pp. 29–34.
Aylward, S.R., Hemminger, B.H., Pisano, E.D., 1998. Mixture modelling for digital mammogram display and analysis. In: International Workshop on Digital Mammography, pp. 305–312.
Ball, J.E., Butler, T.W., Bruce, L.M., 2004. Towards automated segmentation and classification of masses in digital mammograms. In: IEEE Conference on Engineering in Medicine and Biology Society, pp. 1814–1817.
Bårman, H., Granlund, G.H., 1994. Computer aided diagnosis of mammograms using a hierarchical framework. In: International Workshop on Digital Mammography, pp. 271–280.
Basset, L.W., Gold, R.H., 1987. Breast Cancer Detection: Mammograms and Other Methods in Breast Imaging. Grune and Stratton, New York.
Benois, J., Barba, D., 1992. Image segmentation by region-contour cooperation for image coding. In: IAPR International Conference on Pattern Recognition, vol. C, pp. 331–334.
Berns, E.A., Hendrick, R.E., Solari, M., Barke, L., Reddy, D., Wolfman, J., Segal, L., De Leon, P., Benjamin, S., Willis, L., 2006. Digital and screen-film mammography: comparison of image acquisition and interpretation times. Am. J. Roentgenol. 187 (1), 38–41.
Beucher, S., Lenteuejoul, C., 1979. Use of watersheds in contour detection. In: Proceedings of the International Workshop on Image Processing: Real-Time Edge and Motion Detection/Estimation, pp. 2.1–2.12.
Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.
Bird, R.E., Wallace, T.W., Yankaskas, B.C., 1992. Analysis of cancers missed at screening mammography. Radiology 184 (3), 613–617.
Birdwell, R.L., Ikeda, D.M., O'Shaughnessy, K.D., Sickles, E.A., 2001. Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. Radiology 219 (1), 192–202.
Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer, New York.
Bornefalk, H., 2005. Estimation and comparison of CAD system performance in clinical settings. Acad. Radiol. 12, 687–694.
Bornefalk, H., Bornefalk-Hermansson, A., 2005. On the comparison of FROC curves in mammography CAD systems. Med. Phys. 32 (2), 412–417.
Bosch, A., Zisserman, A., Muñoz, X., 2008. Scene classification using a hybrid generative/discriminative approach. IEEE Trans. Pattern Anal. Machine Intell. 30 (4), 712–727.
Bovis, K.J., Singh, S., 2000. Detection of masses in mammograms using texture measures. In: IAPR International Conference on Pattern Recognition, vol. 2, pp. 267–270.
Bruynooghe, M., 2006. Mammographic mass detection using unsupervised clustering in synergy with a parcimonious supervised rule-based classifier. In: Lecturer Notes in Computer Science, vol. 4046, pp. 68–75.
Brzakovic, D., Luo, X.M., Brzakovic, P., 1990. An approach to automated detection of tumors in mammograms. IEEE Trans. Med. Imag. 9 (3), 233–241.
Brzakovic, D., Vujovic, N., Neskovic, M., Brzakovic, P., Fogarty, K., 1994. Mammogram analysis by comparison with previuos screening. In: International Workshop on Digital Mammography, pp. 131–140.
Buseman, S., Mouchawar, J., Calonge, N., Byers, T., 2003. Mammography screening matters for young women with breast carcinoma. Cancer 97 (2), 352–358.
Byng, J.W., Boyd, N.F., Fishell, E., Jong, R.A., Yaffe, M.J., 1996. Automated analysis of mammographic densities. Phys. Med. Biol. 41 (5), 909–923.
Calder, B., Clarke, S., Linnett, L.M., Carmichael, D., 1996. Statistical models for the detection of abnormalities in digital mammography. In: IEE Colloquium Digest Mammography, pp. 6/1–6/6.
Campanini, R., Dongiovanni, D., Iampieri, E., Lanconelli, N., Masotti, M., Palermo, G., Riccardi, A., Roffilli, M., 2004. A novel featureless approach to mass detection in digital mammograms based on support vector machines. Phys. Med. Biol. 49, 961–975.
Cao, A.Z., Song, Q., Yang, X.L., Liu, S., 2004a. Breast mass segmentation on digital mammograms by a combined deterministic annealing method. In: IEEE International Symposium on Biomedical Imaging, vol. 2, pp. 1303–1306.
Cao, A.Z., Song, Q., Yang, X.L., Wang, L., 2004b. Breast mass segmentation based on information theory. In: IAPR International Conference on Pattern Recognition, vol. 3, pp. 758–761.
Cascio, D., Fauci, F., Magro, R., Raso, G., Bellotti, R., De Carlo, F., Tangaro, S., De Nunzio, G., Quarta, M., Forni, G., Lauria, A., Fantacci, M.E., Retico, A., Masala, G.L., Oliva, P., Bagnasco, S., Cheran, S.C., Lopez-Torres, E., 2006. Mammogram segmentation by contour searching and mass lesions classification with neural network. IEEE Trans. Nucl. Sci. 53 (5), 2827–2833.
Catarious, D.M., Baydush, A.H., Floyd, C.E., 2004. Incorporation of an iterative, linear segmentation routine into a mammographic mass CAD system. Med. Phys. 31 (6), 1512–1520.
Catarious, D.M., Baydush, A.H., Floyd, C.E., 2006. Characterization of difference of Gaussian filters in the detection of mammographic regions. Med. Phys. 33 (11), 4104–4114.
Chang, Y.H., Good, W.F., Sumkin, J.H., Zheng, B., Gur, D., 1999. Computerized localization of breast lesions from two views – an experimental comparison of two methods. Invest. Radiol. 34 (9), 585–588.
Chang, Y.H., Zheng, B., Gur, D., 1996a. Computerized identification of suspicious regions for masses in digitized mammograms. Invest. Radiol. 31 (3), 146–153.
Chang, Y.H., Zheng, B., Gur, D., 1996b. Robustness of computerized identification of masses in digitized mammograms – a preliminary assessment. Invest. Radiol. 31 (9), 563–568.
Che, F.N., Fairhust, M.C., Wells, C.P., Hanson, M., 1996. Evaluation of a two-stage model for detection of abnormalities in digital mammograms. In: IEE Colloquium Digest Mammography, pp. 13/1–13/4.
Chen, C.H., Lee, G.G., 1997. On digital mammogram segmentation and microcalcification detection using multiresolution wavelet analysis. Graph. Models Image Process. 59 (5), 349–364.
Chen, P.C., Pavlidis, T., 1979. Segmentation by texture using a co-occurrence matrix and a split-and-merge algorithm. Comput. Graph. Image Process. 10, 172–182.
Cheng, H.D., Cai, X.P., Chen, X.W., Hu, L.M., Lou, X.L., 2003. Computer-aided detection and classification of microcalcifications in mammograms: a survey. Pattern Recogn. 36 (12), 2967–2991.
Cheng, H.D., Cui, M., 2004. Mass lesion detection with a fuzzy neural network. Pattern Recogn. 37, 1189–1200.
Cheng, H.D., Shi, X.J., Min, R., Hu, L.M., Cai, X.P., Du, H.N., 2006. Approaches for automated detection and classification of masses in mammograms. Pattern Recogn. 39 (4), 646–668.
Christoyianni, I., Constantinou, E., Dermatas, E., 2004. Automatic detection of abnormal tissue in bilateral mammograms using neural networks. In: Methods and Applications of Artificial Intelligence, Helenic Conference on AI, pp. 267–275.
Christoyianni, I., Dermatas, E., Kokkinakis, G., 2000. Fast detection of masses in computer-aided mammography. IEEE Signal Process. Mag. 17 (1), 54–64.
Chu, Y., Li, L., Clark, R.A., 2002. Graph-based region growing for mass segmentation in digital mammography. In: Proceedings of SPIE, vol. 4684, pp. 1690–1697.
Comer, M.L., Liu, S., Delp, E.J., 1996. Statistical segmentation of mammograms. In: International Workshop on Digest Mammography, pp. 475–478.
Constantinidis, A.S., Fairhust, M.C., Deravi, F., Hanson, M., Wells, C.P., Chapman-Jones, C., 1999. Evaluating classification strategies for detection of circumscribed masses in digital mammograms. In: International Conference on Image Processing and its Application, pp. 435–439.
Constantinidis, A.S., Fairhust, M.C., Rahman, A.F.R., 2000. Detection of circumscribed masses in digital mammograms using behaviour–knowledge space method. Electron. Lett. 36 (4), 302–303.
Constantinidis, A.S., Fairhust, M.C., Rahman, A.F.R., 2001. A new multi-expert decision combination algorithm and its application to the detection of circumscribed masses in digital mammograms. Pattern Recogn. 34 (8), 1527–1537.
Davies, E.R., 1997. Machine Vision, second ed. Academic Press, London, UK.
De Koning, H.J., Fracheboud, J., Boer, R., Verbeek, A.L., Collette, H.J., Hendriks, J.H.C.L., van Ineveld, B.M., de Bruyn, A.E., van der Maas, P.J., 1995. Nation-wide breast cancer screening in the Netherlands: support for breast cancer mortality reduction. National evaluation team for breast cancer screening. Int. J. Cancer 60 (6), 777–780.
Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum-likelihood from incomplete data via EM algorithm. J. Roy. Stat. Soc. B, 1–38.
Diahi, J.G., Frouge, C., Giron, A., Fertil, B., 1996. Artificial neural networks for detection of breast cancer in mammography. In: International Workshop on Digest Mammography, pp. 329–334.
Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification, second ed. John Wiley and Sons, New York.
Ellis, R.L., Meade, A.A., Mathiason, M.A., Willison, K.M., Logan-Young, W., 2007. Evaluation of computer-aided detection systems in the detection of small invasive breast carcinoma. Radiology 245 (1), 88–94.
Elter, M., Horsch, A., 2009. CADx of mammographic masses and clustered microcalcifications: a review. Med. Phys. 36 (6), 2052–2068.
Eltonsy, N.H., Tourassi, G.D., Elmaghraby, A.S., 2007. A concentric morphology model for the detection of masses in mammography. IEEE Trans. Med. Imag. 26 (6), 880–889.
Eurostat, 2002. Health statistics Atlas on mortality in the European Union. Office for Official Publications of the European Union.
Fauci, F., Bagnasco, S., Bellotti, R., Cascio, D., Cheran, S.C., De Carlo, F., De Nunzio, G., Fantacci, M.E., Forni, G., Lauria, A., Lopez Torrez, E., Magro, R., Masala, G.L., Oliva, P., Quarta, M., Raso, G., Retico, A., Tangaro, S., 2004. Mammogram segmentation by contour searching and massive lesion classification with neural network. In: IEEE Nuclear Science Symposium Conference Record, vol. 5, pp. 2695–2699.
Fauci, F., Raso, G., Magro, R., Forni, G., Lauria, A., Bagnasco, S., Cerello, P., Cheran, S.C., Lopez-Torres, E., Bellotti, R., De Carlo, F., Gargano, G., Tangaro, S., De Mitri, I., De Nunzio, G., Cataldo, R., 2005. A massive lesion detection algorithm in mammography. Phys. Med. 21 (1), 23–30.
Fenton, J.J., Taplin, S.H., Carney, P.A., Abraham, L., Sickles, E.A., D'Orsi, C., Berns, E.A., Cutter, G., Hendrick, R.E., Barlow, W.E., Elmore, J.G., 2007. Influence of computer-aided detection on performance of screening mammography. New Engl. J. Med. 536 (14), 1399–1409.
Filev, P., Hadjiiski, L., Sahiner, B., Chan, H.P., Helvie, M.A., 2005. Comparison of similarity measures for the task of template matching of masses on serial mammograms. Med. Phys. 32 (2), 515–529.
Freer, T.W., Ulissey, M.J., 2001. Screening mammography with computer-aided detection: prospective study of 12860 patients in a community breast center. Radiology 220, 781–786.
Freixenet, J., Muñoz, X., Raba, D., Martí, J., Cufí, X., 2002. Yet another survey on image segmentation: region and boundary information integration. In: European Conference on Computer Vision, vol. III, pp. 408–422.
Freixenet, J., Oliver, A., Lladó, X., Martí, R., Pont, J., Pérez, E., Denton, E., Zwiggelaar, R., 2008. Eigendetection of masses considering false positive reduction and breast density information. Med. Phys. 35 (5), 1840–1853.

Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55 (1), 119–139.
Fu, K.S., Mui, J.K., 1981. A survey on image segmentation. Pattern Recogn. 13, 3–16.
Georgsson, F., 2003. Differential analysis of bilateral mammograms. Int. J. Pattern Recogn. Artif. Intell. 17 (7), 1207–1226.
Giger, M.L., Yin, F.F., Doi, K., Wu, Y., Vyborny, C.J., Schmidt, R.A., Huo, Z., 1992. Computerized detection and characterization of mass lesions in digital mammography. In: IEEE International Conference on Systems, Man, and Cybernetics, pp. 1370–1372.
Giménez, V., Manrique, D., Ríos, J., Vilarrasa, A., 1999. Iterative method for automatic detection of masses in digital mammograms for computer-aided diagnosis. In: Proceedings of SPIE, vol. 3661, pp. 1086–1093.
Good, W.F., Zheng, B., Chang, Y.H., Wang, X.H., Maitz, G., Gur, D., 1999. Multi-image CAD employing features derived from ipsilateral mammographic views. In: Proceedings of SPIE, vol. 3661, pp. 474–485.
Good, W.F., Zheng, B., Chang, Y.H., Wang, Z.H., Maitz, G.S., 2001. Generalized procrustean image deformation for substraction of mammograms. In: Proceedings of SPIE, vol. 3661, pp. 1562–1573.
Goto, M., Morikawa, A., Fujita, H., Hara, T., Endo, T., 1998. Detection of spicules on mammograms based on a multistage pendulum filter. In: International Workshop on Digital Mammography, pp. 135–138.
Groshong, B.R., Kegelmeyer, W.P., 1996. Evaluation of a Hough transform method for circumscribed lesion detection. In: International Workshop on Digital Mammography, pp. 361–366.
Guliato, D., Rangayyan, R.M., Carnielli, W.A., Zuffo, J.A., Desautels, J.E.L., 1998. Segmentation of breast tumors in mammograms by fuzzy region growing. In: IEEE Conference on Engineering in Medicine and Biology Society, vol. 20, pp. 1002–1005.
Guliato, D., Rangayyan, R.M., Carnielli, W.A., Zuffo, J.A., Desautels, J.E.L., 2003. Segmentation of breast tumors in mammograms using fuzzy sets. J. Electron. Imag. 12 (3), 369–378.
Gulsrud, T.O., Engan, K., Hanstveit, T., 2005. Watershed segmentation of detected masses in digital mammograms. In: IEEE Conference on Engineering in Medicine and Biology Society, pp. 3304–3307.
Gupta, R., Undrill, P.E., 1995. The use of texture analysis to identify suspicious masses in mammography. Phys. Med. Biol. 40 (5), 835–855.
Gur, D., Sumkin, J.H., Rockette, H.E., Ganott, M., Hakim, C., Hardesty, L., Poller, W.R., Shah, R., Wallace, L., 2004. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. J. Natl. Cancer Inst. 96 (3), 185–190.
Hachama, M., Desolneux, A., Richard, F., 2006. A probabilistic approach for the simultaneous mammogram registration and abnormality detection. In: Lecturer Notes in Computer Science, vol. 4046, pp. 205–212.
Hadjiiski, L., Sahiner, B., Chan, H.P., Petrick, N., Helvie, M.A., Gurcan, M., 2001a. Analysis of temporal changes of mammographic features: computer-aided classification of malignant and benign breast masses. Med. Phys. 28 (11), 2309–2638.
Hadjiiski, M., Chan, H.P., Sahiner, B., Petrick, N., Helvie, M.A., 2001b. Automated registration of breast lesions in temporal pairs of mammograms for interval change analysis–local affine transformation for improved localization. Med. Phys. 28 (6), 1070–1079.
Hall, F.M., Storella, J.M., Siverstond, D.Z., Wyshak, G., 1988. Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at mammography. Radiology 167 (2), 353–358.
Hassanien, A.E., Ali, J.M., Nobuhara, H., 2004. Detection of spiculated masses in mammograms based on fuzzy image processing. In: Lecturer Notes in Computer Science, vol. 3070, pp. 1002–1007.
Hatanaka, Y., Hara, T., Fujita, H., Kasai, S., Endo, T., Iwase, T., 2001. Development of an automated method for detecting mammographic masses with a partial loss of region. IEEE Trans. Med. Imag. 20 (12), 1209–1214.
Heath, M.D., Bowyer, K.W., 2000. Mass detection by relative image intensity. In: International Workshop on Digital Mammography, pp. 219–225.
Hejazi, M.R., Ho, Y.S., 2005. Automated detection of tumors in mammograms using two segments for classification. In: Lecturer Notes in Computer Science, vol. 3767, pp. 910–921.
Herredsvela, J., Gulsrud, T., Engan, K., 2005. Detection of circumscribed masses in mammograms using morphological segmentation. In: Proceedings of SPIE, vol. 5747, pp. 902–913.
Heywang-Köbrunner, S.H., Dershaw, D.D., Schreer, I., 2001. Diagnostic Breast Imaging. Mammography, sonography, magnetic resonance imaging, and interventional procedures. Thieme, Stuttgart, Germany.
Highnam, R., Kita, Y., Brady, M., Shepstone, B., English, R., 1998. Determining correspondence between views. In: International Workshop on Digital Mammography, pp. 111–118.
Hong, B.W., Brady, M., 2003. A topographic representation for mammogram segmentation. In: Lecturer Notes in Computer Science, vol. 2879, pp. 730–737.
Huo, Z., Giger, M.L., Vyborny, C.J., Bick, U., Lu, P., Wolverton, D.E., Schmidt, R.A., 1995. Analysis of spiculation in the computerized classification of mammographic masses. Med. Phys. 22 (10), 1569–1579.
Jain, A.K., Murty, M.N., Flynn, P.J., 1999. Data clustering: a review. ACM: Computing Surveys 31 (3), 264–323.
Jiang, H., Tiu, W., Yamamoto, S., Iisaku, S., 1998. A method for automatic detection of spicules in mammograms. J. Comput. Aided Diagn. Med. Images 2 (4), 1–8.
Kalman, B.L., Reinus, W.R., Kwasny, S.C., Laine, A., Kotner, L., 1997. Prescreening entire mammograms for masses with artificial neural networks: preliminary results. Acad. Radiol. 4 (6), 405–414.
Karssemeijer, N., 1994. Recognition of stellate lesions in digital mammograms. In: International Workshop on Digital Mammography, pp. 211–219.
Karssemeijer, N., 1998. Automated classification of parenchymal patterns in mammograms. Phys. Med. Biol. 43, 365–378.
Karssemeijer, N., 1999. Local orientation distribution as a function of spatial scale for detection of masses in mammograms. In: Proceedings on Information Processing in Medial Imaging, vol. 1613, pp. 280–293.
Karssemeijer, N., te Brake, G.M., 1996. Detection of stellate distortions in mammograms. IEEE Trans. Med. Imag. 15 (5), 611–619.
Karssemeijer, N., te Brake, G.M., 1998. Combining single view features and asymmetry for detection of mass lesions. In: International Workshop on Digital Mammography, pp. 95–102.
Kasai, S., Kaji, D., Kano, A., Fujita, H., Hara, T., Endo, T., 2002. Mass detection algorithm for digital mammograms based on an adaptive thresholding techniques utilizing multi-resolution processing. In: International Workshop on Digital Mammography, pp. 334–338.
Kegelmeyer, W.P., 1992. Computer detection of stellate lesions in mammograms. In: Proceedings of SPIE, vol. 1660, pp. 446–454.
Kegelmeyer, W.P., Pruneda, J.M., Bourland, P.D., Hillis, A., Riggs, M.W., Nipper, M.L., 1994. Computer-aided mammographic screening for spiculated lesions. Radiology 191 (2), 331–337.
Khan, F., Sarma, A., Sun, Y., Tufts, D., 2002. Mass detection using tolerance intervals and a rank detector. In: IEEE International Symposium on Biomedical Imaging, pp. 185–188.
Khoo, L.A.L., Taylor, P., Given-Wilson, R.M., 2005. Computer-aided detection in the United Kingdom national breast screening programme: prospective study. Radiology 237 (2), 444–449.
Kim, S.J., Moon, W.K., Cho, N., Cha, J.H., Kim, S.M., Im, J.G., 2008. Computer-aided detection in full-field digital mammography: sensitivity and reproducibility in serial examinations. Radiology 246 (1), 71–80.
Kinnard, L., Lo, S.C.B., Makariou, E., Osicka, T., Wang, P., Chouikha, M.F., Freedman, M.T., 2004a. Steepest changes of a probability-based cost function for delineation of mammographic masses: a validation study. Med. Phys. 31 (10), 2796–2810.
Kinnard, L., Lo, S.C.B., Makariou, E., Osicka, T., Wang, P., Freedman, M.T., Chouikha, M.F., 2004b. Likelihood function analysis for segmentation of mammographic masses for various margin groups. In: IEEE International Symposium on Biomedical Imaging, vol. 1, pp. 113–116.
Kinnard, L., Lo, S.C.B., Wang, P., Freedman, M.T., Chouikha, M.F., 2002. Automated segmentation of mammographic masses using fuzzy shadow and maximum-likelihood analysis. In: IEEE International Symposium on Biomedical Imaging, pp. 241–244.
Kita, Y., Highnam, R., Brady, M., 1998. Correspondence between different view breast X-rays using a simulation of breast deformation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 700–707.
Kita, Y., Highnam, R., Brady, M., 2001. Correspondence between different view breast X-rays using curved epipolar lines. Comput. Vis. Image Understanding 83 (1), 38–56.
Kobatake, H., Murakami, M., 1996. Adaptive filter to detect rounded convex regions: Iris filter. In: IAPR International Conference on Pattern Recognition, vol. 2, pp. 340–345.
Kobatake, H., Murakami, M., Takeo, H., Nawano, S., 1999. Computerized detection of malignant tumors on digital mammograms. IEEE Trans. Med. Imag. 18 (5), 369–378.
Kobatake, H., Takeo, H., Nawano, S., 1998. Tumor detection system for full-digital mammography. In: International Workshop on Digital Mammography, pp. 87–94.
Kobatake, H., Yoshinaga, Y., 1996. Detection of spicules on mammogram based on skeleton analysis. IEEE Trans. Med. Imag. 15 (3), 235–245.
Kobatake, H., Yoshinaga, Y., Murakami, M., 1994. Automatic detection of malignant tumors on mammogram. In: IEEE International Conference on Image Processing, vol. 1, pp. 407–410.
Kok-Wiles, S.L., Brady, M., Highman, R., 1998. Comparing mammogram pairs for the detection of lesions. In: International Workshop on Digital Mammography, pp. 103–110.
Kom, G., Tiedeu, A., Kom, M., 2007. Automated detection of masses in mammograms by local adaptive thresholding. Comput. Biol. Med. 37 (1), 37–48.
Kopans, D., 1998. Breast Imaging. Lippincott-Raven, Philadelphia.
Kumar, M.P., Torr, P.H.S., Zisserman, A., 2005. OBJCUT. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 18–25.
Kupinski, M.A., Giger, M.L., 1998. Automated seeded lesion segmentation on digital mammograms. IEEE Trans. Med. Imag. 17 (4), 510–517.
Kuzmiak, C.M., Millnamow, G.A., Qaqish, B., Pisano, E.D., Cole, E.B., Brown, M.E., 2002. Comparison of full-field digital mammography to screen-film mammography with respect to diagnostic accuracy of lesion characterization in breast tissue biopsy specimens. Acad. Radiol. 9, 1378–1382.
Kwok, S.M., Chandrasekhar, R., Attikiouzel, Y., 2002. Adaptation of the Daugman–Downing texture demodulation to highlight circumscribed mass lesions on mammograms. In: International Conference on Digital Signal Processing, pp. 449–452.
Lai, S.M., Li, X., Bischof, W.F., 1989. On techniques for detecting circumscribed masses in mammograms. IEEE Trans. Med. Imag. 8 (4), 377–386.
Laine, A., Huda, W., Chen, D., Harris, J., 1996. Segmentation of masses using continuous scale representations. In: International Workshop on Digital Mammography, pp. 447–450.
Laine, A., Huda, W., Steinbach, B.G., Honeyman, J.C., 1995. Mammographic image processing using wavelet processing techniques. Epidemiol. Rev. 5 (5), 518–523.

Lau, T., Bischof, W.F., 1991. Automated detection of breast tumors using the asymmetry approach. Comput. Biomed. Res. 24 (3), 273–295.
Lee, Y.J., Park, J.M., Park, H.W., 2001. Mammographic mass detection by adaptive thresholding and region growing. Int. J. Imag. Syst. Technol. 11 (5), 340–346.
Leung, J.W.T., Sickles, E.A., 2000. Multiple bilateral masses detected on screening mammography: assessment of need for recall imaging. Am. J. Roentgenol. 175 (1), 23–29.
Li, H., Wang, Y., Liu, K.J.R., Lo, S.C.B., Freedman, M.T., 2001. Computerized radiographic mass detection – part I: lesion site selection by morphological enhancement and contextual segmentation. IEEE Trans. Med. Imag. 20 (4), 289–301.
Li, H.D., Kallergi, M., Clarke, L.P., Jain, V.K., Clark, R.A., 1995. Markov random field for tumor detection in digital mammography. IEEE Trans. Med. Imag. 14 (3), 565–576.
Li, J., Liu, K.J.R., Wang, Y., Lo, S.C.B., 1996a. Morphological filtering and stochastic modeling-based segmentation of masses on mammographic images. In: IEEE Nuclear Science Symposium Conference Record, vol. 3, pp. 1792–1796.
Li, J., Liu, K.J.R., Wang, Y., Lo, S.C.B., 1996b. Nonlinear filtering enhancement and histogram modeling segmentation of masses for digital mammograms. In: IEEE Conference on Engineering in Medicine and Biology Society, pp. 1045–1046.
Li, L., Clark, R.A., Thomas, J.A., 2002. Computer-aided diagnosis of masses with full-field digital mammography. Acad. Radiol. 9 (1), 4–12.
Li, L., Qian, W., Clarke, L.P., 1997. Digital mammography: computer assisted diagnosis method for mass detection with multi-orientation and multiresolution wavelet transform. Acad. Radiol. 4 (11), 724–731.
Li, L., Qian, W., Clarke, L.P., Clark, R.A., Thomas, J.A., 1999. Improving mass detection by adaptive and multiscale processing in digitized mammograms. In: Proceedings of SPIE, vol. 3661, pp. 490–498.
Liu, S., Babbs, C.F., Delp, E.J., 2001. Multiresolution detection of spiculated lesions in digital mammograms. IEEE Trans. Image Process. 10 (6), 874–884.
Lladó, X., Oliver, A., Freixenet, J., Martí, R., Martí, J., 2009. A textural approach for mass false positive reduction in mammography. Comput. Med. Imag. Grap. 33 (6), 415–422.
Lo, S.C.B., Li, H., Wang, Y., Kinnard, L., Freedman, M.T., 2002. A multiple circular path convolution neural network system for detection of mammographic masses. IEEE Trans. Med. Imag. 21 (2), 150–158.
Lobregt, S., Viergever, M.A., 1995. A discrete dynamic contour model. IEEE Trans. Med. Imag. 14, 12–24.
MacQueen, J.B., 1967. Some methods of classification and analysis of multivariate observations. In: Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281–297.
Martí, J., Freixenet, J., Muñoz, X., Oliver, A., 2003. Active region segmentation of mammographic masses based on texture, contour, and shape features. In: Lecturer Notes in Computer Science, vol. 2652, pp. 478–485.
Martí, R., Raba, D., Oliver, A., Zwiggelaar, R., 2006. Mammographic registration: proposal and evaluation of a new approach. In: Lecturer Notes in Computer Science, vol. 4046, pp. 213–220.
Martí, R., Zwiggelaar, R., Rubin, C.M.E., 2001. Tracking mammographic structures over time. In: British Machine Vision Conference, pp. 143–152.
Martí, R., Zwiggelaar, R., Rubin, C.M.E., 2002. Automatic point correspondence and registration based on linear structures. Int. J. Pattern Recogn. Artif. Intell. 16 (3), 331–340.
Matsubara, T., Fujita, H., Hara, T., Kasai, S., Otsuka, O., Hatanaka, Y., Endo, T., 1998. Development of a new algorithm for detection of mammographic masses. In: International Workshop on Digital Mammography, pp. 139–142.
Matsubara, T., Fujita, H., Kasai, S., Goto, M., Tani, Y., Hara, T., Endo, T., 1997. Development of new schemes for detection and analysis of mammographic masses. In: International Conference on Information Systems, pp. 63–66.
McKenzie, P., Alder, M., 1994. Initializing the EM algorithm for use in Gaussian mixture modeling. In: Proceedings on Pattern Recognition in Practice, vol. IV, pp. 91–105.
Méndez, A.J., Souto, M., Tahoces, P.G., Vidal, J.J., 2003. Computer aided diagnosis for breast masses detection on a telemammography system. Comp. Med. Imag. Grap. 27, 497–502.
Méndez, A.J., Tahoces, P.G., Lado, M.J., Souto, M., Vidal, J.J., 1998. Computerized-aided diagnosis: automatic detection of malignant masses in digitized mammograms. Med. Phys. 25 (6), 957–964.
Metz, C.E., 1996. Evaluation of digital mammography by ROC analysis. In: International Workshop on Digital Mammography, pp. 61–68.
Miller, L., Ramsey, N., 1996. The detection of malignant masses by non-linear multiscale analysis. In: International Workshop on Digital Mammography, pp. 335–340.
Morrison, S., Linnett, L.M., 1999. A model based approach to object detection in digital mammography. In: IEEE International Conference on Image Processing, vol. 2, pp. 182–186.
Mousa, R., Munib, Q., Moussa, A., 2005. Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural. Expert Syst. Appl. 28, 713–723.
Mudigonda, N.R., Rangayyan, R.M., Desautels, J.E.L., 2001. Detection of breast masses in mammograms by density slicing and texture flow-field analysis. IEEE Trans. Med. Imag. 20 (12), 1214–1215.
Nakagawa, T., Hara, T., Fujita, H., Iwase, T., Endo, T., Horita, K., 2004. Automated contour extraction of mammographic mass shadow using an improved active contour model. In: International Congress Series, vol. 1268, pp. 882–885.
National Electrical Manufacturers Association, 2006. Digital Imaging and Communications in Medicine (DICOM), third ed. National Electrical Manufacturers Association.
Ng, S.L., Bischof, W.F., 1992. Automated detection and classification of breast tumors. Comput. Biomed. Res. 25 (3), 218–237.
Nishikawa, R.M., Kallergi, M., 2006. Computer-aided detection, in its present form, is not an effective aid for screening mammography. Med. Phys. 33 (4), 811–814.
Obenauer, S., Hermann, K., Grabbe, E., 2003. Dose reduction in full-field digital mammography: an anthropomorphic breast phantom study. Br. J. Radiol. 76 (907), 478–482.
Öktem, V., Jouny, I., 2004. Automatic detection of malignant tumors in mammograms. In: IEEE Conference on Engineering in Medicine and Biology Society, pp. 1770–1773.
Oliver, A., Freixenet, J., Martí, R., Denton, E.R.E., Zwiggelaar, R., 2006. Mammographic mass eigendetection. In: Medical Image Understanding and Analysis, pp. 71–75.
Oliver, A., Freixenet, J., Martí, R., Pont, J., Pérez, E., Denton, E., Zwiggelaar, R., 2008. A novel breast tissue density classification methodology. IEEE Trans. Inform. Technol. Biomed. 12 (1), 55–65.
Özekes, S., Osman, O., Çamurcu, A.Y., 2005. Mammographic mass detection using a mass template. Korean J. Radiol. 6 (4), 221–228.
Pappas, T.N., 1992. An adaptive clustering algorithm for image segmentation. IEEE Trans. Syst. Man Cybern. 40 (4), 901–914.
Paquerault, S., Petrick, N., Chan, H.P., Sahiner, B., Helvie, M.A., 2002. Improvement of computerized mass detection on mammograms: fusion of two-view information. Med. Phys. 29 (2), 238–247.
Parr, T., Astely, S., Boggis, C., 1994. The detection of stellate lesions in digital mammograms. In: International Workshop on Digital Mammography, pp. 231–239.
Petrick, N., Chan, H.P., Sahiner, B., Helvie, M.A., 1999. Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms. Med. Phys. 26 (8), 1642–1654.
Petrick, N., Chan, H.P., Sahiner, B., Wei, D., 1996a. An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection. IEEE Trans. Med. Imag. 15 (1), 59–67.
Petrick, N., Chan, H.P., Wei, D., Sahiner, B., Helvie, M.A., Adler, D.D., 1996b. Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification. Med. Phys. 23 (10), 1685–1696.
Petrick, N., Chan, H.P., Wei, D., Sahiner, B., Wei, D., Helvie, M.A., Goodsit, M.M., Adler, D.D., 1995. Automated detection of breast masses on digital mammograms using adaptive density-weighted contrast-enhancement filtering. In: Proceedings of SPIE, vol. 2434, pp. 590–597.
Petrick, N., Sahiner, B., Chan, H.P., Helvie, M.A., Paquerault, S., Hadjiiski, L.M., 2002. Breast cancer detection: evaluation of a mass-detection algorithm for computer-aided diagnosis – experience in 263 patients. Radiology 224 (1), 217–224.
Pfisterer, R.S., Aghdasi, F., 1999. Hexagonal wavelets for the detection of masses in digitised mammograms. In: Proceedings of SPIE, vol. 3813, pp. 966–977.
Pfisterer, R.S., Aghdasi, F., 2001. Tumor detection in digitized mammograms by image texture analysis. Opt. Eng. 40 (2), 209–216.
Pohlman, S., Powell, K.A., Obuchowski, N.A., Chilcote, W.A., Grundfest-Broniatowksi, S., 1996. Quantitative classification of breast tumors in digitized mammograms. Med. Phys. 23 (8), 1337–1345.
Polakowski, W.E., Cournoyer, D.A., Rogers, S.K., DeSimio, M.P., Ruck, D.W., Hoffmeister, J.W., Raines, R.A., 1997. Computer-aided breast cancer detection and diagnosis of masses using difference of Gaussian and derivative-based feature saliency. IEEE Trans. Med. Imag. 16 (6), 811–819.
Pu, J., Zheng, B., Leader, J.K., Gur, D., 2008. An ellipse-fitting based method for efficient registration of breast masses on two mammographic views. Med. Phys. 35 (2), 487–494.
Qi, H., Snyder, W.E., 1998. Lesion detection and characterization in digital mammography by Bézier histograms. In: IEEE Conference on Engineering in Medicine and Biology Society, vol. 2, pp. 1021–1024.
Qian, W., Li, L., Clarke, L.P., 1999. Image feature extraction for mass detection in digital mammography: influence of wavelet analysis. Med. Phys. 26 (3), 402–408.
Qian, W., Li, L., Clarke, L.P., Mao, F., Clark, R.A., 1998a. Adaptive CAD modules for mass detection in digital mammography. In: IEEE Conference on Engineering in Medicine and Biology Society, vol. 2, pp. 1013–1016.
Qian, W., Li, L., Clarke, L.P., Mao, F., Clark, R.A., Thomas, J., 1998b. A computer assisted diagnostic system for mass detection. In: International Workshop on Digital Mammography, pp. 79–86.
Qian, W., Song, D., Lei, M., Sankar, R., Eikman, E., 2007. Computer-aided mass detection based on ipsilateral multiview mammograms. Acad. Radiol. 14 (5), 530–538.
Qian, W., Sun, X., Song, D., Clark, R.A., 2001. Digital mammography – wavelet transform and Kalman-filtering neural network in mass segmentation and detection. Acad. Radiol. 8 (11), 1074–1082.
Raba, D., Oliver, A., Martí, J., Peracaula, M., Espunya, J., 2005. Breast segmentation with pectoral muscle suppression on digital mammograms. In: Lecturer Notes in Computer Science, vol. 3523, pp. 471–478.
Rangayyan, R.M., Ayres, F.J., Desautels, J.E.L., 2007. A review of computer-aided diagnosis of breast cancer: toward the detection of subtle signs. J. Frankl. Inst. 344 (3–4), 312–348.
Rangayyan, R.M., El-Faramawy, E.N.M., Desautels, J.E.L., Alim, O.A., 1997. Measures of acutance and shape for classification of breast tumors. IEEE Trans. Med. Imag. 16 (6), 799–810.

Richard, F.J.P., Cohen, L.D., 2003. A new image registration technique with free boundary constraints: application to mammography. Comput. Vis. Image Understanding 89, 166–196.
Roberts, L.G., 1965. Machine perception of three-dimensional solids. In: Tippet, J., Berkowitz, D., Clapp, L., Koester, C., Vanderburgh, A. (Eds.), Optical and Electro-Optical Information Processing. MIT Press, Cambridge, MA, pp. 159–197.
Rocha, A., Fu, T., Zhuangzhi, Y., 2000. A logic filter for tumor detection on mammograms. J. Comput. Sci. Technol. 15 (6), 629–632.
Rogova, G.L., Ke, C.C., Acharya, R.S., Stomper, P.C., 1999. Feature choice for detection of cancerous masses by constrained optimization. In: Proceedings of SPIE, vol. 3661, pp. 1440–1447.
Rojas, A., Nandi, A., 2008. Detection of masses in mammograms via statistically based enhancement, multilevel-thresholding segmentation, and region selection. Comput. Med. Imag. Graph. 32 (4), 304–315.
Rose, K., 1998. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. Proc. IEEE 86 (11), 2210–2239.
Sahiner, B., Chan, H.P., Petrick, N., Helvie, M.A., Goodsit, M.M., 1998. Computerized characterization of masses on mammograms: the rubber band straightening transform and texture analysis. Med. Phys. 25 (4), 516–526.
Sahiner, B., Chan, H.P., Petrick, N., Helvie, M.A., Hadjiiski, L.M., 2001a. Improvement of mammographic mass characterization using spiculation measures and morphological features. Med. Phys. 28 (7), 1455–1465.
Sahiner, B., Chan, H.P., Wei, D., Petrick, N., Helvie, M.A., Adler, D.D., Goodsit, M.M., 1996. Image feature selection by a genetic algorithm: application to classification of mass and normal breast tissue. Med. Phys. 23, 1671–1684.
Sahiner, B., Petrick, N., Chan, H.P., Hadjiiski, L.M., Paramagul, C., Helvie, M.A., Gurcan, M.N., 2001b. Computer-aided characterization of mammographic masses: accuracy of mass segmentation and its effects on characterization. IEEE Trans. Med. Imag. 20 (12), 1275–1284.
Sakellaropoulos, F., Skiadopoulos, S., Karahaliou, A., Costaridou, L., Panayiotakis, G., 2006. Using wavelet-based features to identify masses in dense breast parenchyma. In: Lecturer Notes in Computer Science, vol. 4046, pp. 557–564.
Sallam, M., Bowyer, K., 1994. Registering time sequences of mammograms using a two-dimensional image unwarping technique. In: International Workshop on Digital Mammography, pp. 121–130.
Sallam, M., Bowyer, K., 1996. Detecting abnormal densities in mammograms by comparison with previous screenings. In: International Workshop on Digital Mammography, pp. 417–420.
Sameti, M., Ward, R.K., 1996. A fuzzy segmentation algorithm for mammogram partitioning. In: International Workshop on Digital Mammography, pp. 471–474.
Sameti, M., Ward, R.K., Morgan-Parkes, J., Palcic, B., 1997. A method for detection of malignant masses in digitized mammograms using a fuzzy segmentation algorithm. In: IEEE Conference on Engineering in Medicine and Biology Society, pp. 513–516.
Sanjay-Gopal, S., Chan, H.P., Wilson, T., Helvie, M., Petrick, N., Sahiner, B., 1999. A regional registration technique for automated interval change analysis of breast lesions on mammograms. Med. Phys. 26 (12), 2669–2679.
Sheshadri, H.S., Kandaswamy, A., 2005. Detection of breast cancer tumor based on morphological watershed algorithm. Int. J. Graph. Vision Image Process. 5 (5), 17–21.
Shi, J., Sahiner, B., Chan, H.P., Ge, J., Hadjiiski, L.M., Helvie, M.A., Nees, A., Wu, Y.T., Wei, J., Zhou, C., Zhang, Y., Cui, J., 2008. Characterization of mammographic masses based on level set segmentation with new image features and patient information. Med. Phys. 35 (1), 280–290.
Sickles, E.A., 1997. Breast cancer screening outcomes in women ages 40–49: clinical experience with service screening using modern mammography. J. Natl. Cancer Inst.: Monogr. 22, 99–104.
Singh, S., Al-Mansoori, R., 2000. Identification of regions of interest in digital mammograms. J. Intell. Syst. 10 (2), 183–217.
Sivic, J., Russell, B.C., Efros, A.A., Zisserman, A., Freeman, W., 2005. Discovering objects and their location in images. In: IEEE International Conference on Computer Vision, vol. 1, pp. 370–377.
Smith, A.P., 2003. Fundamentals of digital mammograhpy: physics, technology and practical considerations. Radiol. Manag. 25 (5), 18–31.
Stamatakis, E.A., Ricketts, I.W., Cairns, A.Y., Walker, C., Preece, P.E., 1996. Detecting abnormalities on mammograms by bilateral comparison. In: IEE Colloquium Digest Mammography, pp. 12/1–12/4.
Stathaki, T., Constantinides, A.G., 1994. Neural networks and higher order spectra for breast cancer detection. In: IEEE Workshop on Neural Networks and Signal Processing, pp. 473–481.
Suckling, J., Parker, J., Dance, D.R., Astley, S.M., Hutt, I., Boggis, C.R.M., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S.L., Taylor, P., Betal, D., Savage, J., 1994. The Mammographic Image Analysis Society digital mammogram database. In: International Workshop on Digital Mammography, pp. 211–221.
Suliga, M., Deklerck, R., Nyssen, E., 2008. Markov random field-based clustering applied to the segmentation of masses in digital mammograms. Comput. Med. Imag. Graph. 32 (6), 502–512.
Sun, X.J., Qian, W., Song, D.S., 2004. Ipsilateral-mammogram computer-aided detection of breast cancer. Comp. Med. Imag. Graph. 28 (3), 151–158.
Sun, X.J., Qian, W., Song, D.S., Clark, R.A., 2001. Ipsilateral multi-view CAD system for mass detection in digital mammography. In: IEEE Workshop on Mathematical Methods in Biomedical Image Analysis, pp. 19–26.
Székely, N., Tóth, N., Pataki, B., 2006. A hybrid system for detecting masses in mammographic images. IEEE Trans. Instrum. Meas. 55 (3), 944–952.
Tarassenko, L., Hayton, P., Cernaez, N., Brady, M., 1995. Novelty detection for the identification of masses in mammograms. In: IEEE International Conference on Artificial Neural Networks, pp. 442–447.
Taylor, P., Champness, J., Given-Wilson, R., Johnston, K., Potts, H., 2005. Impact of computer-aided detection prompts on the sensitivity and specificity of screening mammography. Health Technol. Assess. 9 (6), 1–58.
te Brake, G.M., Karssemeijer, N., 1998. Comparison of three mass detection methods. In: International Workshop on Digital Mammography, pp. 119–126.
te Brake, G.M., Karssemeijer, N., 1999. Single and multiscale detection of masses in digital mammograms. IEEE Trans. Med. Imag. 18 (7), 628–639.
te Brake, G.M., Karssemeijer, N., 2001. Segmentation of suspicious densities in digital mammograms. Med. Phys. 28 (2), 259–266.
te Brake, G.M., Stoutjesdijk, M.J., Karssemeijer, N., 1999. Discrete dynamic contour model for mass segmentation in digital mammograms. In: Proceedings of SPIE, vol. 3661, pp. 911–919.
Timp, S., Karssemeijer, N., 2004. A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography. Med. Phys. 31 (5), 958–971.
Timp, S., Karssemeijer, N., 2006. Interval change analysis to improve computer aided detection in mammography. Med. Image Anal. 10 (1), 82–95.
Timp, S., van Engeland, S., Karssemeijer, N., 2005. A regional registration method to find corresponding mass lesions in temporal mammogram pairs. Med. Phys. 32 (8), 2629–2638.
Toledo Santos, V., Schiabel, H., Góes, C.E., Benatti, R.H., 2002. A segmentation technique to detect masses in dense breast digitized mammograms. J. Digit. Imag. 15 (1), 210–213.
Torralba, A., Murphy, K.P., Freeman, W.T., 2004. Sharing features: efficient boosting procedures for multiclass object detection. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 762–769.
Tourassi, G.D., Vargas-Vorecek, R., Catarious, D.M., Floyd, C.E., 2003. Computer-assisted detection of mammographic masses: a template matching scheme based on mutual information. Med. Phys. 30 (8), 2123–2130.
Undrill, P., Gupta, R., Henry, S., Downing, M., 1996. Texture analysis and boundary refinement to outline mammography masses. In: IEE Colloquium Digest Mammography, pp. 511–516.
Vacek, P.M., Geller, B.M., Weaver, D.L., Foster, R.S., 2002. Increased mammography use and its impact on earlier breast cancer detection in Vermont. Cancer 94 (8), 2160–2168.
van Engeland, S., Karssemeijer, N., 2006. Exploitation of correspondence between CC and MLO views in computer aided mass detection. In: Lecturer Notes in Computer Science, vol. 4046, pp. 237–242.
van Engeland, S., Karssemeijer, N., 2007. Combining two mammographic projections in a computer aided mass detection method. Med. Phys. 34 (3), 898–905.
Van Engeland, S., Snoeren, P.R., Hendriks, J.H.C.L., Karssemeijer, N., 2003. A comparison of methods for mammogram registration. IEEE Trans. Med. Imag. 22 (11), 1436–1444.
van Engeland, S., Timp, S., Karssemeijer, N., 2006. Finding corresponding regions of interest in mediolateral oblique and craniocaudal mammographic views. Med. Phys. 33 (9), 3203–3212.
Varela, C., Tahoces, P.G., Méndez, A.J., Souto, M., Vidal, J.J., 2007. Computerized detection of breast masses in digitized mammograms. Comput. Biol. Med. 37 (2), 214–226.
Velikova, M., Samulski, M., Lucas, P.J., Karssemeijer, N., 2009. Improved mammographic CAD performance using multi-view information: a Bayesian network frameworks. Phys. Med. Biol. 54 (5), 1131–1147.
Velthuizen, R.P., 2000. Computer diagnosis of mammographic masses. In: Workshop on Applied Imagery Pattern Recognition, pp. 166–172.
Vincent, L., Soille, P., 1991. Watershed in digital spaces: an efficient algorithm based on immersion simulations. IEEE Trans. Pattern Anal. Machine Intell. 13 (6), 583–598.
Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 511–518.
Vujovic, N., Bakic, P.R., Brzakovic, D., 1996. Detection of potentially cancerous signs by mammogram followup. In: International Workshop on Digital Mammography, pp. 421–424.
Vujovic, N., Brzakovic, D., 1997. Establishing the correspondence between control points in pairs of mammographic images. IEEE Trans. Med. Imag. 6 (10), 1388–1399.
Vyborny, C.J., Giger, M.L., 1994. Computer vision and artificial intelligence in mammography. Am. J. Roentgenol. 162 (3), 699–708.
Wai, L.C.C., Brady, M., 2005. Curvilinear structure based mammographic registration. In: Lecturer Notes in Computer Science, vol. 3765, pp. 261–270.
Wang, K., Qin, H., Fisher, P.R., Zhao, W., 2006. Automatic registration of mammograms using texture-based anisotropic features. In: IEEE International Symposium on Biomedical Imaging, pp. 864–867.
Warren, L.J.W., Wood, S.A., D'Orsi, C.J., Feig, S.A., Kopans, D.B., O'Shaughnessy, K.F., Sickles, E., Tabar, L., Vyborny, C.J., Castellino, R., 2000. Potential contribution of computer-aided detection to the sensitivity of screening mammography. Radiology 215 (2), 554–562.
Wei, D., Sahiner, B., Chan, H.P., Petrick, N., 1994. Detection of masses on mammograms using a convolution neural network. In: International Conference on Acoustics, Speech, and Signal Processing, vol. 5, pp. 3483–3486.
Wei, J., Chan, H.P., Sahiner, B., Hadjiiski, L.M., Helvie, M.A., Roubidoux, M.A., Zhou, C., Ge, J., 2006. Dual system approach to computer-aided detection of breast masses on mammograms. Med. Phys. 33 (11), 4157–4168.

Wei, J., Chan, H.P., Sahiner, B., Zhou, C., Hadjiiski, L.M., Roubidoux, M.A., Helvie, M.A., 2009. Computer-aided detection of breast masses on mammograms: dual system approach with two-view analysis. Med. Phys. 36 (10), 4451–4460.

Wei, J., Hadjiiski, L.M., Sahiner, B., Chan, H., Ge, J., Roubidoux, M.A., Helvie, M.A., Zhou, C., Wu, Y.T., Paramagul, C., Zhang, Y., 2007. Computer-aided detection systems for breast masses: comparison of performances on full-field digital mammograms and digitized screen-film mammograms. Acad. Radiol. 14 (6), 659–669.

Wei, J., Sahiner, B., Hadjiiski, L.M., Chan, H.P., Petrick, N., Helvie, M.A., Roubidoux, M.A., Ge, J., Zhou, C., 2005. Computer-aided detection of breast masses on full field digital mammograms. Med. Phys. 32 (9), 2827–2838.

Winsberg, F., Elkin, M., Macy, J., Bordaz, V., Weymouth, W., 1967. Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. Radiology 89 (2), 211–215.

Wirth, M.A., Narhan, J., Gray, D., 2002. Nonrigid mammogram registration using mutual information. In: Proceedings of SPIE, vol. 4684, pp. 95–102.

Woods, K.S., Bowyer, K.W., 1994. Computer detection of stellate lesions. In: International Workshop on Digital Mammography, pp. 221–229.

Wu, Y., Wei, J., Hadjiiski, L.M., Sahiner, B., Zhou, C., Ge, J., Shi, J., Zhang, Y., Chan, H., 2007. Bilateral analysis based false positive reduction for computer-aided mass detection. Med. Phys. 34 (8), 3334–3344.

Xie, M., 2002. A method of tumors detection in digital mammography. In: IEEE International Conference on Communications, Circuits and Systems and West Sino Expositions, vol. 2, pp. 1007–1011.

Xie, M., Ma, Z., 2001. A method of automatic detection of tumors in mammogram. In: Proceedings of SPIE, vol. 4556, pp. 145–153.

Yang, S.K., Moon, W.K., Cho, N., Park, J.S., Cha, J.H., Kim, S.M., Kim, S.J., Im, J.G., 2007. Screening mammography-detected cancers: sensitivity of a computer-aided detection system applied to full-field digital mammograms. Radiology 244 (1), 104–111.

Yin, F.F., Giger, M.L., Doi, K., Vyborny, C.J., Schmidt, R.A., 1991. Computerized detection of masses in digital mammograms: analysis of bilateral subtraction images. Med. Phys. 18 (5), 955–963.

Yin, F.F., Giger, M.L., Doi, K., Vyborny, C.J., Schmidt, R.A., 1994. Computerized detection of masses in digital mammograms: automated alignment of breast images and its effect on bilateral-subtraction technique. Med. Phys. 21 (3), 445–452.

Yin, F.F., Giger, M.L., Vyborny, C.J., Doi, K., Schmidt, R.A., 1993. Comparison of bilateral-substraction and single-image processing techniques in the computerized detection of mammographic masses. Invest. Radiol. 28 (6), 473–481.

Yin, L., Deshpande, S., Chang, J.K., 2003. Automatic lesion/tumor detection using intelligent mesh-based active contour. In: IEEE International Conference on Tools with Artificial Intelligence, pp. 390–397.

Yoon, H.J., Zheng, B., Sahiner, B., Chakraborty, D.P., 2007. Evaluating computer-aided detection algorithms. Med. Phys. 34 (6), 2024–2038.

Youssry, N., Abou-Chadi, F.E.Z., El-Sayad, A.M., 2003. Early detection of masses in digitized mammograms using texture features and neuro-fuzzy model. In: IEEE Conference on Information Technology in Applied Biomedicine, pp. 226–229.

Yuan, Y., Giger, M.L., Li, H., Suzuki, K., Sennett, C., 2007. A dual-stage method for lesion segmentation on digital mammograms. Med. Phys. 34 (11), 4180–4193.

Zhang, H., Foo, S.W., 2006. Computer aided detection of breast masses from digitized mammograms. IEICE Trans. Inform. Syst. E89D (6), 1955–1961.

Zhang, H., Foo, S.W., Krishnan, S.M., Thng, C.H., 2004. Automated breast masses segmentation in digitized mammograms. In: IEEE International Workshop in Biomedical Circuits and System, pp. S2.2–1–S2.2–4.

Zhang, M., Giger, M.L., Vyborny, C.J., Doi, K., 1996. Mammographic texture analysis for the detection of spiculated lesions. In: International Workshop on Digital Mammography, pp. 347–350.

Zheng, B., Chang, Y.H., Gur, D., 1995a. Computerized detection of masses in digitized mammograms: comparison of single-image segmentation and bilateral-image substraction. Acad. Radiol. 2 (12), 1056–1061.

Zheng, B., Chang, Y.H., Gur, D., 1995b. Computerized detection of masses in digitized mammograms using single-image segmentation and a multilayer topographic feature analysis. Acad. Radiol. 2 (11), 959–966.

Zheng, B., Chang, Y.H., Gur, D., 1996. Mass detection in digitized mammograms using two independent computer-assisted diagnosis schemes. Am. J. Roentgenol. 167 (6), 1421–1424.

Zheng, B., Chang, Y.H., Wang, X.H., Good, W.F., Gur, D., 1999a. Feature selection for computerized mass detection in digitized mammograms by using a genetic algorithm. Acad. Radiol. 6 (6), 327–332.

Zheng, B., Good, W.F., Armfield, D.R., Cohen, C., Hertzberg, T., Sumkin, J.H., Gur, D., 2003a. Performance change of mammographic CAD schemes optimized with most-recent and prior image databases. Acad. Radiol. 10 (3), 283–288.

Zheng, B., Hardesty, L.A., Poller, W.R., Sumkin, J.H., Golla, S., 2003b. Computer-aided detection in mammography: a reproducibility assessment – initial experience. Radiology 228 (1), 58–62.

Zheng, B., Leader, J.K., Abrams, G.S., Lu, A.H., Wallace, L.P., Maitz, G.S., Gur, D., 2006. Multiview-based computer-aided detection scheme for breast masses. Med. Phys. 33 (9), 3135–3143.

Zheng, L., Chan, A.K., 2001. An artificial intelligent algorithm for tumor detection in screening mammogram. IEEE Trans. Med. Imag. 20 (7), 559–567.

Zheng, L., Chan, A.K., McCord, G., Wu, S., Liu, J.S., 1999b. Detection of cancerous masses for screening mammography using DWT based multiresolution Markov Random Field. J. Digit. Imag. 12 (1–2), 18–23.

Zhou, C., Chan, H.P., Petrick, N., Helvie, M.A., Goodsitt, M.M., Sahiner, B., Hadjiiski, L.M., 2001. Computerized image analysis: estimation of breast density on mammograms. Med. Phys. 28 (6), 1056–1069.

Zhu, S.C., Yuille, A., 1996. Region competition: unifying snakes, region growing, and Bayes/MDL for multi-band image segmentation. IEEE Trans. Pattern Anal. Machine Intell. 18 (9), 884–900.

Zouras, W.K., Giger, M.L., Lu, P., Wolverton, D.E., Vyborny, C.J., Doi, K., 1996. Investigation of a temporal subtraction scheme for computerized detection of breast masses in mammograms. In: International Workshop on Digital Mammography, pp. 411–415.

Zucker, S.W., 1976. Region growing: childhood and adolescence. Comput. Graph. Image Process. 5, 382–399.

Zwiggelaar, R., Astley, S.M., Taylor, C.J., 1998. Detecting the central mass of a spiculated lesion using scale-orientation signatures. In: International Workshop on Digital Mammography, pp. 63–70.

Zwiggelaar, R., Parr, T.C., Schumm, J.E., Hutt, I.W., Taylor, C.J., Astley, S.M., Boggis, C.R.M., 1999a. Model-based detection of spiculated lesions in mammograms. Med. Image Anal. 3 (1), 39–62.

Zwiggelaar, R., Taylor, C.J., Rubin, C.M.E., 1999. Detection of the central mass of spiculated lesions – signature normalisation and model data aspects. In: Lecturer Notes in Computer Science, vol. 1613, pp. 406–411.